

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 7, 2004, 19:46:08 ; Search time 26569 Seconds
(without alignments)
7119.161 Million cell updates/sec
Title: US-10-788-793-1
Perfect score: 4364
Sequence: 1 ccactgggtttcttcaaggga.....aaaaaaaaaaaaaaaaaaaaa 4364

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_ey:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4364	100.0	4364	10	AB055759	AB055759 Rattus no
2	3636.4	83.3	3755	10	D87257	D87257 Rattus norv
3	3031.4	69.5	4641	9	BSM807322	BX647178 Homo sapi
4	3025.6	69.3	4892	6	AR339574	AR339574 Sequence
5	2987.2	68.5	4247	9	AB086011	AB086011 Homo sapi
6	2804.8	64.3	227874	2	AC130850	AC130850 Rattus no
7	2804.8	64.3	238626	2	AC108529	AC108529 Rattus no
8	2797.4	64.1	4051	9	AB033101	AB033101 Homo sapi
9	2760.2	63.2	4065	6	AX780844	AX780844 Sequence
10	2641	60.5	4793	9	BSM803316	AL832009 Homo sapi
11	2556.6	58.6	139753	2	AC121534	AC121534 Mus muscu
12	2556.6	58.6	191609	2	AC140247	AC140247 Mus muscu
13	2443.2	56.0	3559	9	AK027705	AK027705 Homo sapi
14	2431.2	55.7	3558	6	AX877347	AX877347 Sequence
15	2431.2	55.7	3558	6	BD156611	BD156611 Primer fo
16	2080	47.7	44768	9	AL589649	AL589649 Human DNA
17	1729.4	39.6	2594	6	AX877018	AX877018 Sequence
18	1729.4	39.6	2594	6	BD156441	BD156441 Primer fo
19	1729.4	39.6	2594	9	AK001570	AK001570 Homo sapi
20	1191.2	27.3	1798	6	AX834513	AX834513 Sequence
21	1191.2	27.3	1798	9	AK097021	AK097021 Homo sapi
22	790.2	18.1	3415	6	AX776296	AX776296 Sequence
23	788.6	18.1	3416	6	AX776298	AX776298 Sequence
24	788.6	18.1	3430	6	AX776294	AX776294 Sequence
25	777.6	17.8	3998	6	AX776292	AX776292 Sequence
26	777.6	17.8	3998	9	AF329092	AF329092 Homo sapi
27	630.8	14.5	3229	9	BC027860	BC027860 Homo sapi
28	630	14.4	117000	9	AC069222	AC069222 Homo sapi
29	628.8	14.4	221863	2	AC095323	AC095323 Rattus no
30	628.4	14.4	166992	2	AC022883	AC022883 Homo sapi
31	625.6	14.3	3303	10	BC062166	BC062166 Mus muscu
32	622.2	14.3	236538	2	AC144776	AC144776 Mus muscu
33	581.6	13.3	844	6	AX868362	AX868362 Sequence
34	581.6	13.3	844	6	BD148424	BD148424 Primer fo
35	565	12.9	3069	6	AX834002	AX834002 Sequence
36	565	12.9	3069	9	AK096195	AK096195 Homo sapi
37	560	12.8	808	6	AX868849	AX868849 Sequence
38	560	12.8	808	6	BD148911	BD148911 Primer fo
39	541.6	12.4	3025	6	AX375016	AX375016 Sequence
40	541.6	12.4	3025	9	HSU53445	U53445 Human Ovari
41	496.6	11.4	2355	6	AX776306	AX776306 Sequence
42	365.4	8.4	1241	6	AX053651	AX053651 Sequence
43	292	6.7	1050	6	AX776290	AX776290 Sequence
44	285.6	6.5	215199	2	AC127889	AC127889 Rattus no
45	265.4	6.1	795	6	AX776288	AX776288 Sequence

ALIGNMENTS

RESULT 1
AB055759
LOCUS
DEFINITION Rattus norvegicus mRNA for filamin-interacting protein L-FILIP,
4364 bp mRNA linear ROD 16-AUG-2002
complete cds.
ACCESSION AB055759
VERSION AB055759.1 GI:21392398
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1
AUTHORS Nagano,T., Yoneda,T., Hatanaka,Y., Kubota,C., Murakami,F. and

Db 1321 ACAGCAAGGAACCTTAGACTAGAAGTGGAGAAAGCTGCAGAAGAGGATGTCTGAGCTGGAGA 1380

QY 1381 AGCTGGAGGAAGCGTTTCAGCCGGAGTTAAGTCGGAATGCACCCAGCTCCATCTCTGAAACCTGG 1440

Db 1381 AGCTGGAGGAAGCGTTTCAGCCGGAGTTAAGTCGGAATGCACCCAGCTCCATCTGAAACCTGG 1440

QY 1441 AGAAGGAGAAAGAACCTTAACCAAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGAG 1500

Db 1441 AGAAGGAGAAAGAACCTTAACCAAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGAG 1500

QY 1501 TTAAGAAACTCGAATGCTCCGAGAGTAGACTGGAGAAAGCCGAGTTAAGCCTCAAAGATG 1560

Db 1501 TTAAGAAACTCGAATGCTCCGAGAGTAGACTGGAGAAAGCCGAGTTAAGCCTCAAAGATG 1560

QY 1561 ACCTTACAAAGCTGAAAGTCTTTCACCTGTGATGCTGGTGGATGAGAGGAAATAATATGATGG 1620

Db 1561 ACCTTACAAAGCTGAAAGTCTTTCACCTGTGATGCTGGTGGATGAGAGGAAATAATATGATGG 1620

QY 1621 AGAAATAAAGCAAGAGAGAGGAAAGTGGATGGGTTGAATAAAACTTTAAGGTGGAGC 1680

Db 1621 AGAAATAAAGCAAGAGAGAGGAAAGTGGATGGGTTGAATAAAACTTTAAGGTGGAGC 1680

QY 1681 AGGGAAGTCAATGGATGTGACGGAAAGCTAAATCGAGGAAAGCAAGAGCTTTTAAAAAC 1740

Db 1681 AGGGAAGTCAATGGATGTGACGGAAAGCTAAATCGAGGAAAGCAAGAGCTTTTAAAAAC 1740

QY 1741 TCAAACTCTGAAATGGAGGAAAGGAGTACAGTCTGACAAAGGAGGGATGAGCTGATGG 1800

Db 1741 TCAAACTCTGAAATGGAGGAAAGGAGTACAGTCTGACAAAGGAGGGATGAGCTGATGG 1800

QY 1801 GTAAACTGAGGAGCGAAGAAAGGTCCTGTGAACTGAGCTGCAGTGTAGACTTACTAA 1860

Db 1801 GTAAACTGAGGAGCGAAGAAAGGTCCTGTGAACTGAGCTGCAGTGTAGACTTACTAA 1860

QY 1861 AGAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGGGAAATAAAACCGAGGTAGGTCTGTGCA 1920

Db 1861 AGAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGGGAAATAAAACCGAGGTAGGTCTGTGCA 1920

QY 1921 AGGGGTCTGAGTTTACCTGCCCCGGAAGACATAAGATCAGAGAACTAACCGTTGAAATCG 1980

Db 1921 AGGGGTCTGAGTTTACCTGCCCCGGAAGACATAAGATCAGAGAACTAACCGTTGAAATCG 1980

QY 1981 AGAGACTGAAGAAACGGCTCCAGCAGTTGGAGTGGTGGAGGGGACTTGTATGAAGACCG 2040

Db 1981 AGAGACTGAAGAAACGGCTCCAGCAGTTGGAGTGGTGGAGGGGACTTGTATGAAGACCG 2040

QY 2041 AGGACGAATATGACCAAGTGGAGCAGAAAGTTTCAAGACCGAGCAGGATAAGCCAAACTTCC 2100

Db 2041 AGGACGAATATGACCAAGTGGAGCAGAAAGTTTCAAGACCGAGCAGGATAAGCCAAACTTCC 2100

QY 2101 TCTCCCAGCAGCTCGAGGAAATCAAAACATGGCCCAAGCACAAGCCATAGAGAAAG 2160

Db 2101 TCTCCCAGCAGCTCGAGGAAATCAAAACATGGCCCAAGCACAAGCCATAGAGAAAG 2160

QY 2161 GGGAGGCCGTGAGCCAGGAAGCCGAACCTGCGACACAGGTTTCGGCTGGAGAGGCTAAAA 2220

Db 2161 GGGAGGCCGTGAGCCAGGAAGCCGAACCTGCGACACAGGTTTCGGCTGGAGAGGCTAAAA 2220

QY 2221 GTCGTGATTTACAGSCCGAGGTGCAGGCTCTCAAGGAGAAGATCCACGAGCTGATGAACA 2280

Db 2221 GTCGTGATTTACAGSCCGAGGTGCAGGCTCTCAAGGAGAAGATCCACGAGCTGATGAACA 2280

QY 2281 AGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTATTCGGTCTTCAGCAAGATTTATGG 2340

Db 2281 AGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTATTCGGTCTTCAGCAAGATTTATGG 2340

QY 2341 AAGAAGAAACTAAGAAACAAGAACATGGGGAGGAGGTCTCAATCTGACCAAGGAGCTAG 2400

Db 2341 AAGAAGAAACTAAGAAACAAGAACATGGGGAGGAGGTCTCAATCTGACCAAGGAGCTAG 2400

QY 2401 AGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAAGGATGGTGG 2460

Db 2401 AGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAAGGATGGTGG 2460

QY 2461 ACGTGCCTGTGGCCTCCACTGGGTGCAGACCGAGGGGTGTGCGGGATGCTGCGGAGG 2520

Db 2461 ACGTGCCTGTGGCCTCCACTGGGTGCAGACCGAGGGGTGTGCGGGATGCTGCGGAGG 2520

QY 2521 AGGAGACCCCGGCTGTGTTTCAATTCGCAAAATCCTTCCAGGAGGAAAAATCAGATCATGAGTA 2580

Db 2521 AGGAGACCCCGGCTGTGTTTCAATTCGCAAAATCCTTCCAGGAGGAAAAATCAGATCATGAGTA 2580

QY 2581 ATCTTCGACAGGTAGGCCTGAAGAAACCCATGGAACCGTCTCCGTCTCTCGACAGGTATC 2640

Db 2581 ATCTTCGACAGGTAGGCCTGAAGAAACCCATGGAACCGTCTCCGTCTCTCGACAGGTATC 2640

QY 2641 CCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTGGATTCTTGGATGAGAAAAAGAG 2700

Db 2641 CCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTGGATTCTTGGATGAGAAAAAGAG 2700

QY 2701 AAAACGGTCTTCCACTCCGAGGAGAAAGGGCCCCAGGCCAAACCAAGGTGCAGGGCACC 2760

Db 2701 AAAACGGTCTTCCACTCCGAGGAGAAAGGGCCCCAGGCCAAACCAAGGTGCAGGGCACC 2760

QY 2761 CCGGGGAGCTGGTCTTAGCACCAAAAGCAGGSCCAGCCCCCTACACATCCCGTGTACACCCAG 2820

Db 2761 CCGGGGAGCTGGTCTTAGCACCAAAAGCAGGSCCAGCCCCCTACACATCCCGTGTACACCCAG 2820

QY 2821 ATCATGAGAACAGCACTGCCACCTTGAGATCACAAGCCCCACATCTGAGAGATTTTTTCT 2880

Db 2821 ATCATGAGAACAGCACTGCCACCTTGAGATCACAAGCCCCACATCTGAGAGATTTTTTCT 2880

QY 2881 CTAGTACCACCGTCATTCTACCTTAGGCAACCAGAAAACCAAGAATAACCATTTATTTCCAT 2940

Db 2881 CTAGTACCACCGTCATTCTACCTTAGGCAACCAGAAAACCAAGAATAACCATTTATTTCCAT 2940

QY 2941 CACCCAAATGTGTCGAAAAAGCCCCAAAAGTGAGATCCTACTCTCGGCCCAGAAACGAG 3000

Db 2941 CACCCAAATGTGTCGAAAAAGCCCCAAAAGTGAGATCCTACTCTCGGCCCAGAAACGAG 3000

QY 3001 CCATGTCCCCTGTACGATTACTACTATTTCCAGAGAGAGAGCCCCGGAAGGTGGAAGGA 3060

Db 3001 CCATGTCCCCTGTACGATTACTACTATTTCCAGAGAGAGAGCCCCGGAAGGTGGAAGGA 3060

QY 3061 GCGCCTTTGCCGACAGGCCCTGCATCCCCCATCCAAATCATGACGGTGTCAACATCTGCAG 3120

Db 3061 GCGCCTTTGCCGACAGGCCCTGCATCCCCCATCCAAATCATGACGGTGTCAACATCTGCAG 3120

QY 3121 CTCCCACTGAAATCGCTGTCTCTCCTGAAATCTCAGGAAGTGCCTATGGGAAGGACTATCC 3180

Db 3121 CTCCCACTGAAATCGCTGTCTCTCCTGAAATCTCAGGAAGTGCCTATGGGAAGGACTATCC 3180

QY 3181 TCAAAGTCACCCCGGAAAAACAAACTGTTCCAGCCCCCGTGCAGGAAGTACAACTCCAATG 3240

Db 3181 TCAAAGTCACCCCGGAAAAACAAACTGTTCCAGCCCCCGTGCAGGAAGTACAACTCCAATG 3240

QY 3241 CTAATATCATCAACACGGAAGACAATAAAATTCACATTCACCTGGGTTCTCAGTTTAAAGC 3300

Db 3241 CTAATATCATCAACACGGAAGACAATAAAATTCACATTCACCTGGGTTCTCAGTTTAAAGC 3300

QY 3301 GATCTCCTGGGCTGCCGCTGAAGCGGTAGCCCCAGTTTATCACCGTCCGGSCCTGTCAACG 3360

Db 3301 GATCTCCTGGGCTGCCGCTGAAGCGGTAGCCCCAGTTTATCACCGTCCGGSCCTGTCAACG 3360

QY 3361 TGACAGCGGAGAGGAGGTTTCTACAGGCACAGTCTTTCGCTCTCCAGGAACCACTCT 3420

Db 3361 TGACAGCGGAGAGGAGGTTTCTACAGGCACAGTCTTTCGCTCTCCAGGAACCACTCT 3420

QY 3421 CTTCAAGACCCCGTGTAGCAAAAGTGACGACACTATAACTATAACCCCGGTCAACCGT 3480

Db 3421 CTTCAAGACCCCGTGTAGCAAAAGTGACGACACTATAACTATAACCCCGGTCAACCGT 3480

QY 3481 CATCCACACGAGGAACCCCAATCAGTGTGAGGAAAGATGGGTCTCTCAGCGGCTTACCC 3540

Db 3481 CATCCACACGAGGAACCCCAATCAGTGTGAGGAAAGATGGGTCTCTCAGCGGCTTACCC 3540

QY	3541	CCACCCGCAATTCCTATGTCAAAAGGTATGAAAGCTGGAAGCCAGTAGTGGCAGCCTCAG	3600
Db	3541		
QY	3601	GAGCAGGAAATCTGACCRAATTCAGCCTCGAGCTGAGACTCAGTCTATGAAAAATAGAGC	3660
Db	3601		
QY	3661	TGAAGAAATCTGCAGCCAGCAGCACTGCCTCTCTTGGAGGGGGGAAGGCTGAGGGCAGT	3720
Db	3661		
QY	3721	GGCTAAGGGGGTATGTTGTAAGGATGCTACTGCTGCAGTGGGAAACAACCTTCCTCTGTG	3780
Db	3721		
QY	3781	CCAACCTTTCCTTGTACTACTAATTTAAGTTTAAATATATCTTGTTTATAAAAATAACCAT	3840
Db	3781		
QY	3841	TTAATAGCCATGCACCCCTCCCATTTTGTGCACTGTTCATCGAGGGGAATAGAAAT	3900
Db	3841		
QY	3901	TAATTAGCAGAAATTCGTGTTTGTGTAATGTTCTGTGTAAGATGTTGGTCCAGTTTCAGTTT	3960
Db	3901		
QY	3961	TACTTCTAGCATGTGGCCCATTCAGGTAGCTCACGAGTTGTGAAGCCCTCAATATCGT	4020
Db	3961		
QY	4021	CACCGGAGAGATTTGAGACCAACATTACATATGCTCCCAAGGCTGGTCCCAATTTCC	4080
Db	4021		
QY	4081	TAATTGTAAGCCAACTTTAAATAGACTCAGTCTCTGTGATTTTTCCTCAAAAAAAAATA	4140
Db	4081		
QY	4141	TTTTGAAATAGGACAGATTTAACAGTTGTCTATTTTGCACTATCAAGCCATGAGTTTGAAT	4200
Db	4141		
QY	4201	ATATGGGTTATAAGAAAAGAAATACTTTCAGAGCTATCACAGGGTCTCTAAAACTTTTCGAA	4260
Db	4201		
QY	4261	AAACAAAGCCCTTAATATGACCTCAGGAAACAATTTGAAACATGAAATAAAATGGAATG	4320
Db	4261		
QY	4321	AACGTGTGGAATCTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA	4364
Db	4321		
RESULT 2			
LOCUS	D87257	Rattus norvegicus mRNA for filamin-interacting protein S-FILIP,	3755 bp mRNA linear ROD 16-AUG-2002
DEFINITION	D87257	complete cds.	
ACCESSION	D87257		
VERSION	D87257.1	GI:21392396	
KEYWORDS			
SOURCE		Rattus norvegicus (Norway rat)	
ORGANISM		Rattus norvegicus	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;	
		Rattus.	
REFERENCE	1		
AUTHORS		Nagano, T., Yoneda, T., Hatanaka, Y., Kubota, C., Murakami, F. and Sato, M.	

TITLE	Filamin A-interacting protein (FILIP) regulates cortical cell migration out of the ventricular zone			
JOURNAL	Nat. Cell Biol. 4 (7), 495-501 (2002)			
MEDLINE	22100428			
PUBMED	12055638			
REFERENCE	2 (bases 1 to 3755)			
AUTHORS	Yoneda,T. and Sato,M.			
TITLE	Direct Submission			
JOURNAL	Submitted (20-AUG-1996) Makoto Sato, Fukui Medical University, Department of Anatomy (2); 23-3 Shimoaizuki, Matsuoka, Fukui, 910-1193, Japan (E-mail:makosato@fmsrsa.fukui-med.ac.jp, Tel:81-776-61-8303(ex.2205), Fax:81-776-61-8155)			
COMMENT	Sequence updated (18-May-1999).			
FEATURES	Location/Qualifiers			
source	1..3755			
	/organism="Rattus norvegicus"			
	/mol_type="mRNA"			
	/strain="Wistar"			
	/db_xref="taxon:10116"			
	/clone="yoneda sato clone ki"			
	/tissue_type="forebrain"			
	/clone_lib="library of T.Yoneda"			
	/dev_stage="embryonic day 11.5"			
gene	1..3755			
	/gene="S-FILIP"			
CDS	237..3134			
	/gene="S-FILIP"			
	/codon_start=1			
	/product="filamin-interacting protein S-FILIP"			
	/protein_id="BAC00852.1"			
	/db_xref="GI:21392397"			
	/translation="MLVDERQMHIEQLGQSQKVQD LTKLRBEEEEKLVATYKSKED ROKLKLKLEVD FEHKASRFQSEHEEMNAKLANQESHNRQLRLKVLGLSQRIEELEETNK SLQAAEEELQELREKIAKGECNSSLMAEVSRLKRVLEMGKDEEITKTQAQRELK KKLQEEHHSKELRLEVEKLQKRMSELEKLEAFSRSKSECTQLHLNLEKEKNLTJDKL LNELEVKS RVKELECSERLEKAE LSKDDLTKLKSFTVMLVDKRNMMMEKIKQEBER KVDGLNKNFKVEQKVMVDTEKLI EESKLLKLKSEMEEXEYSLTKERDELMGKLRSE EERSCELSVDLLKKRLDGLIEEVEREINRGRSKGSEFTCPEDNKRILETLIERLK KRLQQLVEVGDLMKTD EYDQLEQKFRTEQDKANFLSQQL EEEKHQMAKHKAIEKGE AVSQEAELRHRFRLEEA KSRDLOAEVOALKEKIH ELMNKEDQLSLOVDYSVLQQRFM EEBTKNKNMGREVLNLTKELELSK RYSRALRPSGNGRRWVDVFASTGVQTEAVCGDA ABEETPAVFI RKSFOEENHIMSNLRQVGLK KPMERSSVLD RYPPAANELTMRKSWIPW MRKRENGPSTPOEKGPRNQAGHPGELVLAPKQGQPLHIRVTPDHENSTALEITSP TSEEFFSSTVITLGNQPRITII PSNVMSQKPKSADPTLGPERAMSPVTTITISR EKSPEGGSAPADRPA SPQIMTVSTSAAPTEIAVSPESQEVPMGRTILKVPEKQTV PAPVRKYNSNANIITTEDNKIHLG SQFKRSPGPAAEVSPVITVRPNVTAEKEVS TGTVLRSPRNHLSSRPGASKVTSTITITPVTTSSRTGTQTSVSGQDGSQRPTPTPIPM SKGMKAGKPVVAASGAGNLTKFPRAETQSMKIELKKSAASTASLGGGKG"			
ORIGIN				
Query Match	83.3%;	Score 3636.4;	DB 10;	Length 3755;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 3637;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	697	AGAGTTGAAAAAGCTCTTGAACAAAGAAAAAGCTTACCAAGCCCGCAAGAAAAAGGAAA	756	
Db	118	AAAGTTGAAAAAGCTCTTGGACAAAGAAAAAGCTTACCAAGCCCGCAAGAAAAAGGAAA	177	
Qy	757	ACGCTAAGCGGCTCAACAAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCA	816	
Db	178	ACGCTAAGCGGCTCAACAAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCA	237	
Qy	817	TGTTGGTGACGAGAGGCAGATGCACATCGAGCAACTGGGCCCTGCAGAGTCAGAAAGTCC	876	
Db	238	TGTTGGTGACGAGAGGCAGATGCACATCGAGCAACTGGGCCCTGCAGAGTCAGAAAGTCC	297	
Qy	877	AGGACCTCACTCAGAACTGAGGAGGAGGAAGAAAACTCAAAGCGGTCACTTACAAAT	936	
Db	298	AGGACCTCACTCAGAACTGAGGAGGAGGAAGAAAACTCAAAGCGGTCACTTACAAAT	357	
Qy	937	CCAAGGAAGACCGCCAGAAGCTGCTCAAGTTAGAACTGGACTTCGAAACACAAAGGCCTCGA	996	
Db	358	CCAAGGAAGACCGCCAGAAGCTGCTCAAGTTAGAACTGGACTTCGAAACACAAAGGCCTCGA	417	

QY	997	GGTTTTCCAGGAGCACGAAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCACAACC	1056
Db	418	GGTTTTCCAGGAGCACGAAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCACAACC	477
QY	1057	GGCAACTTCGACTCAAACCTGGTTGGCTTATCGCAAAAGGATTGAGGAGCTGGAAGACCA	1116
Db	478	GGCAACTTCGACTCAAACCTGGTTGGCTTATCGCAAAAGGATTGAGGAGCTGGAAGACCA	537
QY	1117	ATAAAGCCTTCAGAAAGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAATTGCCAAAG	1176
Db	538	ATAAAGCCTTCAGAAAGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAATTGCCAAAG	597
QY	1177	GGGAATGTGGAACCTCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCTGCTTG	1236
Db	598	GGGAATGTGGAACCTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCTGCTTG	657
QY	1237	AGATGGAGGGCAAGGATGAAGAGATCACGAAGACCGAGGCCAGTGC	1296
Db	658	AGATGGAGGGCAAGGATGAAGAGATCACGAAGACCGAGGCCAGTGC	717
QY	1297	AGAAGCTCCAAGAGGAAGAACACACAGCAAGGAACCTTAGACTAGAAGTGGAGAGCTGC	1356
Db	718	AGAAGCTCCAAGAGGAAGAACACACAGCAAGGAACCTTAGACTAGAAGTGGAGAGCTGC	777
QY	1357	AGAAGAGGATGCTGAGCTGGAGAAGCTGGAGGAAGCGTTAGCGCGGAGTAAGTCGGAAT	1416
Db	778	AGAAGAGGATGCTGAGCTGGAGAAGCTGGAGGAAGCGTTAGCGCGGAGTAAGTCGGAAT	837
QY	1417	GCACCCAGCTCCATCTGAACCTGGAGAAGGAGAGAAACCTAACCAAGACCTGCTGAACG	1476
Db	838	GCACCCAGCTCCATCTGAACCTGGAGAAGGAGAGAAACCTAACCAAGACCTGCTGAACG	897
QY	1477	AGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGA	1536
Db	898	AGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGA	957
QY	1537	AGCCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAGTCTTCACTGTGATCTGG	1596
Db	958	AGCCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAGTCTTCACTGTGATCTGG	1017
QY	1597	TGGATGAGAGGAAAAATATGATGGAGAAAAATAAGCAAGAAGAGAGGAAAGTGGATGGT	1656
Db	1018	TGGATGAGAGGAAAAATATGATGGAGAAAAATAAGCAAGAAGAGAGGAAAGTGGATGGT	1077
QY	1657	TGAATAAAAACTTTAAGGTGGAGCAGGGAAAAAGTCATGGATGTGACGGAAAAAGCTAAATCG	1716
Db	1078	TGAATAAAAACTTTAAGGTGGAGCAGGGAAAAAGTCATGGATGTGACGGAAAAAGCTAAATCG	1137
QY	1717	AGGAAAGCAAGAGCTTTTAAAACTCAAATCTGAAATGGAGGAAAAAGGATACAGTCTGA	1776
Db	1138	AGGAAAGCAAGAGCTTTTAAAACTCAAATCTGAAATGGAGGAAAAAGGATACAGTCTGA	1197
QY	1777	CAAAGGAGGGATGAGCTGATGGGTAAACTGAGGAGCGAAGAAGGTCCTGTGAAC	1836
Db	1198	CAAAGGAGGGATGAGCTGATGGGTAAACTGAGGAGCGAAGAAGGTCCTGTGAAC	1257
QY	1837	TGAGCTGCAGTGTAGACTTACTAAAGAACCGGCTTGATGGCATAGAGGAGGTAGAAAGGG	1896
Db	1258	TGAGCTGCAGTGTAGACTTACTAAAGAACCGGCTTGATGGCATAGAGGAGGTAGAAAGGG	1317
QY	1897	AAATAAACCGAGGTAGGTGCTGCAAGGGGTCTGAGTTCACTGCCCGGAAGACAATAAGA	1956
Db	1318	AAATAAACCGAGGTAGGTGCTGCAAGGGGTCTGAGTTCACTGCCCGGAAGACAATAAGA	1377
QY	1957	TCAGAGAACTAACGCTTGAAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTGG	2016
Db	1378	TCAGAGAACTAACGCTTGAAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTGG	1437
QY	2017	TGGAGGGGACTTGATGAAGACCGGAGCGAATATGACCAGTTGGAGCGAGAAGTTCAGAA	2076
Db	1438	TGGAGGGGACTTGATGAAGACCGGAGCGAATATGACCAGTTGGAGCGAGAAGTTCAGAA	1497

QY	2077	CCGAGCAGGATAAGGCAAACTTCTCTCTCCAGCAGCTCGAGGAAATCAAAACACCAAAATGG	2136
Db	1498	CCGAGCAGGATAAGGCAAACTTCTCTCTCCAGCAGCTCGAGGAAATCAAAACACCAAAATGG	1557
QY	2137	CAAAGCACAAAGCCATAGAGAAAAGGGAGGCCGTGAGCCAGGAAGCCGAACTGCGACACA	2196
Db	1558	CAAAGCACAAAGCCATAGAGAAAAGGGAGGCCGTGAGCCAGGAAGCCGAACTGCGACACA	1617
QY	2197	GGTTTCGGCTGGAGGAGCTAAAAGTCTGTATTTACAGGCCGAGGTGCAGGCTCTCAAGG	2256
Db	1618	GGTTTCGGCTGGAGGAGCTAAAAGTCTGTATTTACAGGCCGAGGTGCAGGCTCTCAAGG	1677
QY	2257	AGAAGATCCACGAGCTGATGAA CAAGGAAGACCCAGCTGTCTCAGCTCCAAGTCGACTATT	2316
Db	1678	AGAAGATCCACGAGCTGATGAA CAAGGAAGACCCAGCTGTCTCAGCTCCAAGTCGACTATT	1737
QY	2317	CGGTCCTTCAGCAAAAGATTATGGAAGAAGAACTAAGAAACAAGAACATGGGAGGGAGG	2376
Db	1738	CGGTCCTTCAGCAAAAGATTATGGAAGAAGAACTAAGAAACAAGAACATGGGAGGGAGG	1797
QY	2377	TCCTCAATCTGACCAAGGAGCTAGAGCTTTCAAAGCGCTACAGCCGAGCTCTCAGGCCGA	2436
Db	1798	TCCTCAATCTGACCAAGGAGCTAGAGCTTTCAAAGCGCTACAGCCGAGCTCTCAGGCCGA	1857
QY	2437	GTGGAAACGGCCGAAGGATGGTGGACGTGCCTGTGTGGCTTCCACTGGGTGCAGACCGAGG	2496
Db	1858	GTGGAAACGGCCGAAGGATGGTGGACGTGCCTGTGTGGCTTCCACTGGGTGCAGACCGAGG	1917
QY	2497	CGGTGTGCGGGATGCTGCGGAGGAGGAGACCCCGGCTGTGTTTTCATTCGCAAAATCCTTCC	2556
Db	1918	CGGTGTGCGGGATGCTGCGGAGGAGGAGACCCCGGCTGTGTTTTCATTCGCAAAATCCTTCC	1977
QY	2557	AGGAGGAAAAATCACATCATGAGTAATCTTCGACAGGTAGGCCTGAAGAAAAACCATGGAAC	2616
Db	1978	AGGAGGAAAAATCACATCATGAGTAATCTTCGACAGGTAGGCCTGAAGAAAAACCATGGAAC	2037
QY	2617	GGTCTCGGTCTCGACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTT	2676
Db	2038	GGTCTCGGTCTCGACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTT	2097
QY	2677	GGATTCTTGATGAGAAAAAGAGAAAAACGGTCTTCCACTCCGAGGAGAAAAAGGGCCCA	2736
Db	2098	GGATTCTTGATGAGAAAAAGAGAAAAACGGTCTTCCACTCCGAGGAGAAAAAGGGCCCA	2157
QY	2737	GGCCAAACCCAGGTGCAGGGCACCCCGGGAGCTGGTCTTAGCACCAAGCAGGGCCAGC	2796
Db	2158	GGCCAAACCCAGGTGCAGGGCACCCCGGGAGCTGGTCTTAGCACCAAGCAGGGCCAGC	2217
QY	2797	CCCTACACATCCGTGTGACACCAGATCATGAGAACAGCAGCTGCCACCCCTGAGATCACAA	2856
Db	2218	CCCTACACATCCGTGTGACACCAGATCATGAGAACAGCAGCTGCCACCCCTGAGATCACAA	2277
QY	2857	GCCCCACATCTGAAGAGTTTTTCTCTAGTACCACCGTCATTCCTACCTTAGGCAACCAGA	2916
Db	2278	GCCCCACATCTGAAGAGTTTTTCTCTAGTACCACCGTCATTCCTACCTTAGGCAACCAGA	2337
QY	2917	AACCAAGATAACCATTTATTCATCACCCCAATGTCTGTACGATTACTACTATTTCCAGAG	2976
Db	2338	AACCAAGATAACCATTTATTCATCACCCCAATGTCTGTACGATTACTACTATTTCCAGAG	2397
QY	2977	ATCCTACTCTCGGCCCAGAAACGAGCCATGTCCCTGTACGATTACTACTATTTCCAGAG	3036
Db	2398	ATCCTACTCTCGGCCCAGAAACGAGCCATGTCCCTGTACGATTACTACTATTTCCAGAG	2457
QY	3037	AGAAGAGCCCCGGAAGGTGGAAGGAGCGGCTTTGCCGACAGGCCTGCATCCCCCATCCAAA	3096
Db	2458	AGAAGAGCCCCGGAAGGTGGAAGGAGCGGCTTTGCCGACAGGCCTGCATCCCCCATCCAAA	2517
QY	3097	TCATGACGGTGTCAACATCTGACGCTCCACTGAAATCGCTGTCTCTCTGAATCTCAGG	3156
Db	2518	TCATGACGGTGTCAACATCTGACGCTCCACTGAAATCGCTGTCTCTCTGAATCTCAGG	2577
QY	3157	AAGTGCCTATGGGAAGGACTATCTCTCAAAGTCAACCCCGGAAAAACAAACTGTTCCAGCCC	3216

Db 2578 ||||| AAGTGCCTATGGGAAGGACTATCCTCAAAGTCACCCCGGAAAAACAACACTGTCTCCAGCCC 2637

Qy 3217 CCGTGC GGAAGTACAACCTCCAATGCTAATATCATCACCGGAAGACAATAAAATTCACA 3276

Db 2638 CCGTGC GGAAGTACAACCTCCAATGCTAATATCATCACCGGAAGACAATAAAATTCACA 2697

Qy 3277 TTCACCTGGGTTCTCAGTTTAAAGCGATCTCTCTGGGCCTGCCGCTGAAGGCGTGAGCCCGAG 3336

Db 2698 TTCACCTGGGTTCTCAGTTTAAAGCGATCTCTCTGGGCCTGCCGCTGAAGGCGTGAGCCCGAG 2757

Qy 3337 TTATCACCGTCCGGCCTGTCAACGTGACAGCGGAGAAAGGAGTTTCTACAGGCACAGTCC 3396

Db 2758 TTATCACCGTCCGGCCTGTCAACGTGACAGCGGAGAAAGGAGTTTCTACAGGCACAGTCC 2817

Qy 3397 TTCGCTCTCCAGGAACCACTCTCTTCAAGACCCGGTGCTAGCAAAAGTGACCACTA 3456

Db 2818 TTCGCTCTCCAGGAACCACTCTCTTCAAGACCCGGTGCTAGCAAAAGTGACCACTA 2877

Qy 3457 TAACTATAACCCCGGTACAAACGTATCCACACGAGGAAACCCCAATCAGTGTCAAGGACAAG 3516

Db 2878 TAACTATAACCCCGGTACAAACGTATCCACACGAGGAAACCCCAATCAGTGTCAAGGACAAG 2937

Qy 3517 ATGGGTATCTCAGCGGCTACCCCAACCCGCAATCTATGTCAAAAGGTATGAAAGCTG 3576

Db 2938 ATGGGTATCTCAGCGGCTACCCCAACCCGCAATCTATGTCAAAAGGTATGAAAGCTG 2997

Qy 3577 GAAAGCCAGTAGTGGCAGCTCAGGAGCAGGAAATCTGACCAAAATTCACGCTCGAGCTG 3636

Db 2998 GAAAGCCAGTAGTGGCAGCTCAGGAGCAGGAAATCTGACCAAAATTCACGCTCGAGCTG 3057

Qy 3637 AGACTCAGTCTATGAAATPAGAGCTGAAGAAATCTGCAGCAGCAGCACTGCCTCTCTTG 3696

Db 3058 AGACTCAGTCTATGAAATPAGAGCTGAAGAAATCTGCAGCAGCAGCACTGCCTCTCTTG 3117

Qy 3697 GAGGGGGAAGGGCTGAGGGCAGTGGCTAAGGGGGTATGTTGTAAGGATGCTACTGCTGC 3756

Db 3118 GAGGGGGAAGGGCTGAGGGCAGTGGCTAAGGGGGTATGTTGTAAGGATGCTACTGCTGC 3177

Qy 3757 AGTGGAACAACAACCTTCCTCTGTGCCAACCCCTTTCCTTGCTACTACTAAATTAAGTTTAA 3816

Db 3178 AGTGGAACAACAACCTTCCTCTGTGCCAACCCCTTTCCTTGCTACTACTAAATTAAGTTTAA 3237

Qy 3817 ATATCTTGTTTATAAATAAACCATTTAATAGCCATGACCCCCCTCCCATTTTGTGCAATC 3876

Db 3238 ATATCTTGTTTATAAATAAACCATTTAATAGCCATGACCCCCCTCCCATTTTGTGCAATC 3297

Qy 3877 TGTTCATGTCAGGGGAATAGAATTAATPAGCAGAAATTTCTGTTTGCTGAATGTTCTGTT 3936

Db 3298 TGTTCATGTCAGGGGAATAGAATTAATPAGCAGAAATTTCTGTTTGCTGAATGTTCTGTT 3357

Qy 3937 GAAGATGTTGGTCCAGTTCAGTTTACTTCTAGCATGTGGCCCCCATTCAGAGGTAGCTCAC 3996

Db 3358 GAAGATGTTGGTCCAGTTCAGTTTACTTCTAGCATGTGGCCCCCATTCAGAGGTAGCTCAC 3417

Qy 3997 GAGTTGTGAAGCCCTCAATATCGTCACCGGAGAGATTTGAGGACCAACATACATATGCTC 4056

Db 3418 GAGTTGTGAAGCCCTCAATATCGTCACCGGAGAGATTTGAGGACCAACATACATATGCTC 3477

Qy 4057 CCAAAGGCTGGCTCCCAATTTTCCTAATTTGAAGCCAACTTTAATAGACTCAGTTCCTGTG 4116

Db 3478 CCAAAGGCTGGCTCCCAATTTTCCTAATTTGTAAGCCAACTTTAATAGACTCAGTTCCTGTG 3537

Qy 4117 ATTTTTCCTCCAAAAAATAATTTTGAAATAGGACAGATTTTAACAGTTGTCAATTTT 4176

Db 3538 ATTTTTCCTCCAAAAAATAATTTTGAAATAGGACAGATTTTAACAGTTGTCAATTTT 3597

Qy 4177 GCACATCAAGCCCATGAGTTTGATATATGGGTTTATAGAAAAAGAAATACTTTTCAGAGCTAT 4236

Db 3598 GCACATCAAGCCCATGAGTTTGATATATGGGTTTATAGAAAAAGAAATACTTTTCAGAGCTAT 3657

Qy 4237 CACAGGCTCTCTAAACTTTTGGAAAAACAAAGCCCTTAATATGACCTCAGGAACAATT 4296

Db 3658 CACAGGCTCTCTAAACTTTTGGAAAAACAAGGCCCTTAATATGACCTCAGAAAACAATT 3717

Qy 4297 TGAACATGAAATAAAATGGAAATGAACCTGTGGAAATCTT 4334

Db 3718 TGAACATGAAATAAAATGGAAATGAACCTGTGGAAATCTT 3755

RESULT 3

HSM807322

LOCUS HSM807322 4641 bp mRNA linear PRI 30-AUG-2003

DEFINITION Homo sapiens mRNA; cDNA DKFZp686M12223 (from clone DKFZp686M12223).

ACCESSION BX647178

VERSION BX647178.1 GI:34366206

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 4641)

AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.

CONSRMT The German Human cDNA Consortium

TITLE Direct Submission

JOURNAL Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY

COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.

This clone (DKFZp686M12223) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cDNA/.

FEATURES

source Location/Qualifiers

1..4641

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFZp686M12223"

/tissue_type="human small intestine"

/clone_lib="686 (synonym: hlcc3). Vector pSport1_sfi; host DH10B; sites SfiIA + SfiIB"

/dev_stage="adult"

polyA_signal 4592..4597

polyA_site 4620

ORIGIN

Query Match 69.5%; Score 3031.4; DB 9; Length 4641;

Best Local Similarity 82.5%; Pred. No. 0;

Matches 3600; Conservative 0; Mismatches 736; Indels 27; Gaps 10;

Qy 1 CCACCTGGGTTCTTCRAAGGGATAAACCCAGCGGC-GAAAGAACACACCATTTGGTT-AAGGAG 58

Db 298 CCACCTGGGTTCTTCAAAGGATAAACTACCTACATAGAGGACATACCTTTGGTTAAAGGAG 357

Qy 59 TCGACAACACAGGTGGGAATGAGATCAGCAAAATCAAGGTGGAGAAAAGTTTCATCTAACGGGCA 118

Db 358 CTGCCGTCAAGTGGGAATGAGATCTCGAAACCAAGGTGGTGAAGTGATCTGATGGGCA 417

Qy 119 TGTCTCTGCCCCAAGTCTCTCCATCATCAGCAGTGAATGGTGGTAAGGGCCCCCTCAGAAGA 178

Db 418 TATCTCTGTCCCAAGCCCTCCATCATCGGCAATGCTGGTGA AAAAAGTCTCTCAGAAGA 477

Qy 179 TGCAAAAAAGAACAAAG---GCCAATCGGAAGGAGGAGGATGTCTATGGCTTCCGGAACCTAT 235

Db 478 TGCAAAAAAGAGAGAAATCAATATAGGAAGGAGGATGTCTATGGCCTCAGGAACTGT 537

Qy 236 CAAAGGCCACCTCAAAACCATCTGGAGAAAGTGAGAAAAAGACTAAGAGTCTGTGGAGTT 295

Db 538 CAAAGCACACCTTAAAAACATCTGGAGAATGTGAACGAAAAAACTAAGAAATCCCTGGAGTT 597

Qy 296 ATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATATGGAAGGGGAGTTGCAGGCTCGAGA 355

Db 598 ATCCAAAGAAGACCTCATCCAACTACTCAGTATAATGGAAGGGGAGTTGCAGGCCAGAGA 657
QY 356 AGATGTATCCACATGCTGAGGACAGAGAAAACCAAGCCCGAGGTTCTGGAGGCACACTA 415
Db 658 AGATGTATCCACATGCTGAAGACAGAGAAAACCAAGCCTGAGGTTCTGGAGGCTCATTA 717
QY 416 TGGATCTGCAGAACTTGAGAAAAGTGCTTCGGGTCCTGCAACCGAGATGCCATCCTTGCTCA 475
Db 718 CGGGTCTCGGAGCCAGAGAAAAGTGCTGCGGGTCCTGCACCGAGATGCCATTCCTGCCCCA 777
QY 476 AGAGAACTCCATAGGAGAAAGACGTCTATGAGAAAACCTATCTCAGAGCTGGACAGACTGGA 535
Db 778 GGAGAAATCCATAGGAGAAAGATGTCTATGAGAAAACCGATTTTCAGAGCTGGACAGACTTGA 837
QY 536 GGAAAAGCAGAAAGGAGACGTACCGCCGCATGCTAGAGCAGCTGCTGCTGGCTGAGAAGTG 595
Db 838 GGAAAACACAGAAAAGAACTTACCGCGCATGCTAGAGCAGCTGTTGCTGGCCGAGAAGTG 897
QY 596 TCACAGGCGCACCGTGTACGAGCTGGAGAAACGAGAAAGCACAAGCACACTGACTACATGAA 655
Db 898 TCATAGGCGCACCGTATACGAGTTAGAGAACGAGAAAGCATAAACACACTGACTACATGAA 957
QY 656 CAAGAGCGACGACTTCACCAACCTGCTGGAGCAGAGCGGAGAGAGTTGAAAAGCTCCT 715
Db 958 CAAGAGCGACGACTTCACCAACCTGCTGGAGCAGAGCGGGAGAGGTTAAAAAGCTCCT 1017
QY 716 TGAACAAGAAAAAGCTTTACCAAGCCCGCAAGAAAAGGAAAACGCTAAGCGGCTCAACAA 775
Db 1018 TGAACAAGAAAAGGCTTATCAAGCCCGCAAGAAAAGGAAAATGCTAAACGACTCAATAA 1077
QY 776 ACTTCGAGATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCATGTTGGTGACGAGAGGCA 835
Db 1078 ACTAAGAGATGAGCTTGTAAACTCAAACTCCTTTGCACCTCATGCTGGTGATGAAAAGACA 1137
QY 836 GATGCACATCGAGCAACTGGGCCTGCAGAGTCAGAAAAGTCCAGGACCTCACTCAGAAGCT 895
Db 1138 AATGCACATTGAACAACCTTGGCCTGCAAGCCAGAAAAGTACAGGATCTTACTCAGAAGCT 1197
QY 896 GAGGAGGAGGAAGAAAACCTCAAAGCGGTCACTTACAAATCCAAGGAAGACCGCCAGAA 955
Db 1198 GAGGGAAGAAAGAGAAAGCTCAAAGCCATTACTTCCAAATCCAAGAAAGACAGACAGAA 1257
QY 956 GCTGCTCAAAGTTAGAGTGGAATTCGAACACAAAGGCCTCGAGGTTTCCAGGAGCACGA 1015
Db 1258 ATTGCTCAAAGTTAGAGTGGAATTTGAACACAAAGGCTTCGAGGTTTTCTCAAGAGCATGA 1317
QY 1016 AGAGATGAACGCCAAATTTGGCGAATCAAGAAATCTCAAAACCGGCAACTTCGACTCAAACT 1075
Db 1318 AGAGATGAACGCTAAACTGGCTAATCAAGAGTCTCAAAATAGGCAACTTAGACTCAAGCT 1377
QY 1076 GGTGGCTTATCGCAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAGGC 1135
Db 1378 GGTGGCTTAAACCCAAAGAAATCGAGGAGCTAGAAGAGACCAA CAAAAATCTGCAGAAGGC 1437
QY 1136 AGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAATTTGCCAAAGGGGAATGTGGAACCTCCAG 1195
Db 1438 AGAGGAAGAACTTCAAGAATTAAGAGATAAAATTGCCAAAGGAGAATGTGGAGACTCTAG 1497
QY 1196 TCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTCTTGAGATGGAGGGCAAGGATGA 1255
Db 1498 CCTCATGGCAGAAAGTGGAAAATCTTCGAAAGCGTGTGCTTGAAATGGAAGTAAAGATGA 1557
QY 1256 AGAGATCAGGAAGACCGAGGCCAGTGCAGGAGCTGAAGAAGAAAGCTTCAAGAGGGAAGA 1315
Db 1558 GGAGATCACTAAACTGAATCCAGTGTAGGGAATGAGGAAGAGCTGCAAGAGGGAAGA 1617
QY 1316 ACACCACAGCAAGGAACTTAGACTAGAAAGTGGAGAAGCTGCAGAAGAGGATGTCTGAGCT 1375
Db 1618 ACACCATAGTAAGGAGCTCAGACTTGAAGTTGAGAAGCTACAGAAGAGAGATGTCTGAACT 1677
QY 1376 GGAGAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAA 1435

Db 1678 AGAGAAATTTGGAAGAAGCATTATTAGCAAGAGTAAATCTGAGTGCACCCAGCTACATTTAA 1737
QY 1436 CCTGGAGAAGGAGAAACCTTAACCAAAGACCTTGCTGTAACGAGCTGGAGGTGTCAGAG 1495
Db 1738 TCTGGAGAAAGAAAAGAACTTAACCAAAGACCTTGCTAAATGAATTTGGAGGTGGTCAAGAG 1797
QY 1496 TCGAGTTAAAGAACTCGAAATGCTCCGAGAGTAGACTGGAGAAAGCCGAGTTAAGCTCAA 1555
Db 1798 TCGAGTTAAAGAAATTTGGAATGTTCTGAAAAGTAGATTGGAAAAGGCTGAAATTAAGCCTAAA 1857
QY 1556 AGATGACCTTACAAAAGCTGAAAGTCTTTCACCTGTGATGCTGGTGATGAGAGGAAAATAT 1615
Db 1858 AGATGATCTTACCAAAGTTGAAGTCATTTACCGTGATGCTGGTTGATGAAAGSRAAAATAT 1917
QY 1616 GATGGAGAAAATAAAAGCAAAGAGAGGAAAAGTGGATGGTTGAAATAAAAACTTTAAAGGT 1675
Db 1918 GATGGAAAAATAAAACAAAGAGAGAAAAGTGGATGGACTCAATAAAAAATTTTAAGGT 1977
QY 1676 GGAGCAGGGAAAAAGTTCATGGATGTGACGGGAAAAGCTAATCGAGGAAAAGCAAGAGCTTTT 1735
Db 1978 GGAACAAGGAAAAGTTATGGATGTAACTGAAAAAACTAATTGAAGAAAAGTAAAGAACTTTT 2037
QY 1736 AAAACTCAAATCTGAAATGGAGGAAAAGGAGTACAGTCTGACAAAAGGAGAGGATGAGCT 1795
Db 2038 AAAACTAAAATCTGAAATGGAGGAAAAGTATACAACTTGACAAAGAGAAAAGATGAGTT 2097
QY 1796 GATGGGTAACCTGAGGAGCGAAGAAAGGAGTCTGCTGTGAAGTCTGAGCTGCAGTGTAGACTT 1855
Db 2098 GATAGGCAAAATTGAAAAAGTGAGGAAGAAAAATCCTCTGAATTAAGCTGCAGTGTGACTT 2157
QY 1856 ACTAAAGAAAGCGGCTTGATGGCATAGAGGAGGTAGAAAAGGGAATAAAACCGAGGTAGGTC 1915
Db 2158 ACTAAAGAAGAGACTTGATGGTATAGAGGAAGTGGAAAAGAGAAATAACAAGAGGAAGGTC 2217
QY 1916 GTGCAAGGGGCTCTGAGTTACCTGCCCGGAAGACAATAAGATCAGAGAACTTAACGCTTGA 1975
Db 2218 ACGAAAAGGCTCTGAGCTCACCTGCCCGGAAGATATAAGATTAAGGAACTAACACTTGA 2277
QY 1976 AATCGAGAGACTGAAGAAAACGGCTCCAGCAGTTGGAGGTGGAGGGGACTTGATGAA 2035
Db 2278 AATTGAGAGACTGAAGAAAACGTCTCCAACAAATTGGAAGTGGTCGAAGGGGATTTGATGAA 2337
QY 2036 GACCGAGGACGAATATGACCAGTTGGAGCAGAAGTTTCAGAA CCGAGCAGGATAAGGCCAAA 2095
Db 2338 GACAGAAGATGAATATGATCAGCTGGAACACAGAAAATTTAGAACTGAGCAGGATAAAGGCTAA 2397
QY 2096 CTTCTCTCCAGCAGCTCGAGGAAATCAAAACACCAAAATGGCCAAGCACAAGCCATAGA 2155
Db 2398 CTTCTCTCTCAACAACCTAGAGGAGATCAAGCACCAAAATTGCCAAGAATAAAGCAATAGA 2457
QY 2156 GAAAGGGGAGCGCTGAGCCAGGAAGCCGAACTCGACACAGGTTTCGGCTCGAGGAGGC 2215
Db 2458 GAAAGGTGAGGTTGTGAGCCAGGAAGCTGAACTGAGACACAGATTTTCGGTTGGAAGAAGC 2517
QY 2216 TAAAAGTCGTGATTTACAGGCCGAGGTGAGGCTCTCAAGGAGAAGATCCACGAGCTGAT 2275
Db 2518 TAAAAGTCGAGACTTAAAAGCCGAAAGTACAAGCTCTTAAAGAGAAGATTCACGAATTAAT 2577
QY 2276 GAACAAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTATTTCGGTCTTCAGCAAAAGATT 2335
Db 2578 GAACAAAGAGATCAGCTTCTCAGCTCCAGGTAGATTTATCTGTACTTCAACAAAGATT 2637
QY 2336 TATGGAAGAGAAAACCTAAGAACAAAGAACATGCGGAGGGAGGTCCTCAATCTGACCAAGGA 2395
Db 2638 TATGGAAGAGAAAATAAGAACAAAACATGCGGCAGGAGGTTCTCAATCTGACCAAGA 2697
QY 2396 GCTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGCCCGAAGGAT 2455
Db 2698 GTTGGAGCTTTCCAAGCGCTACAGCAGAGCTCTTAGGCCACGTGTGAATGGAAGAAGAT 2757
QY 2456 GGTGGACGTGCTGTGGCCTCCAAGTGGGTGCAGACCGAGGCGGTGTGCGGGGATGCTGC 2515
Db 2758 GGTGGATGTTCTGTGACGTCAACTGGAGTCCAAACTGATGCAGTCAGCGGTGAAGCAGC 2817

QY	2516	GGAGGAGGAGACCCCGGCTGTGTTTCATTTCGCAAACTCTTCCAGGAGGAGAAATCACATCAT	2576
Db	2818	AGAGGAAGAAAACGCAGCTGTATTTCATACGGAATCTTCCAGGAAGAAATCATATTAT	2877
QY	2576	GAGTAATCTTCGACAGGTAGGCCTGAAGAAAACCCATGAAACGGTCCTCGGTCTCGACAG	2635
Db	2878	GAGTAATCTTCGGCAGGTGGGATTGAAGAAAACCCGTGAAAGATCTTCTGTCTTAGACAG	2937
QY	2636	GTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTTGGATTCTCTGGATGAGAA	2695
Db	2938	GTATCTCCAGCAGCAAAATGAGCTCACTATGAGAAAGTCTTTGGATTCCATGGATGAGAA	2997
QY	2696	AAGAGAAAACGGTCTTCCACTCCGCAGGAGAAAGGGCCCCAGGCCAAACAGGGTGCAGG	2755
Db	2998	GAGGGAACCGGCCCTCCATCACTCAGGAGAAAGGGCCCCGAAACAAATCCAGTCCAGG	3057
QY	2756	GCACCCCGGGAGCTGGTCTTAGCACCAAGCAGGGCCAGCCCCCTACATCCCGTGTGAC	2815
Db	3058	GCACCCAGGAGGTAGTCTCTTTCACCAAGCAGGCCAGCCCCCTGCATATTCGAGTGC	3117
QY	2816	ACCAGATCATGAGAAACAGCACTGGCCACCTGGAGATCAAGCCCCACATCTGAAGAGTT	2875
Db	3118	ACCAGACCAAGAAACAGCACTGCGACTTTGGAGATAACAAGCCCGACATCTGAAGAATT	3177
QY	2876	TTTTCTTAGTACCACCGTCACTTCTTACCTTAGGCAACCAAGAAACCAAGAATAACCATTTAT	2935
Db	3178	TTTTTCTAGTACCACTGTCTTCTTACCTTAGGGAATCAGAAACCAAGAATAACCATTTAT	3237
QY	2936	TCCATCACCCAATGTCTGTGCGAAAAGCCCCAAAAGTGCAGATCCTACTCTCGGCCCAGA	2995
Db	3238	TCCATCACCAACGGTTATGCCTCAAAAACAAAAAGTGGAGATACTACTCTTGGCCCCAGA	3297
QY	2996	ACGAGCCATGTCCCCTGTACAGATTACTACTATTTCCAGAGAGAAGAGCCCCGGAAGTGG	3055
Db	3298	ACGAGCCATGTCCCCAGTCACAATTACTACTATTTCCAGAGAGAAGACTCCAGAAAGTGG	3357
QY	3056	AAGGAGCGCCTTTGCCGACAGGCCTGCATCCCCCATCCAAATCATGACGGTGTCAACATC	3115
Db	3358	AAGAGCGCATTTGCAGACAGGCCCCACATCCCTATTTCAGATAATGACGGTGTCTACATC	3417
QY	3116	TGCAGTCCCACACTGAAATCGCTGTCTCTCCTGAAATCTCAGGAAGTGCCTATGGGAAGGAC	3175
Db	3418	AGCAGCACGAGTGTGAGATTGACGTTTCTCCGGAATCCCAGGAATGCCCATGGGACGGAC	3477
QY	3176	TATCTCAAAGTCAACCCCGGAAAAACAAACTGTTCCAGCCCGCTGCGGAAGTACAACATC	3235
Db	3478	AATCTCAAAGTCAACCCAGAAAAACAGGTGTTTCCAACTCCAGTACGGAATACAACATC	3537
QY	3236	CAATGCTAATATCATCACCACGGAAGACAATAAAATTCAATTCACTGGGTCTCAGTT	3295
Db	3538	CAATGCCAATATCATTAACACAGAGGACAATAAAATTCAATTCACTTAGGGTCTCAGTT	3597
QY	3296	TAAGCGATCTCCTGGCCTCCGCTGAAGSGGTGAGCCCAAGTTATCACCGTCCGGCCTGT	3355
Db	3598	TAAACGGTCCCCCTGGACTTCAGGTGAAGGAGTCACTCCAGTTATTACTGTCCGACCAGT	3657
QY	3356	CAACGTGACAGCGGAAGAGGAGTTTCTACAGGCACAGTCTCGTCTCCAGGAACCA	3415
Db	3658	AAACGTGACAGCCGAAAAGGAGGTTTCCACCGGCACTGCTCTCGCTCTCCAGGAATCA	3717
QY	3416	CCTCTCTTCAAGACCCCGGTGCTAGCAAAAGTGACCAACTATACTAATACCCCGGTCA	3475
Db	3718	CCTCTCTCAGCGCTGTGTCAGCAAGCAAGTGACGAGCACTATCACCAATAACACCGGTCA	3777
QY	3476	AACGTCATCCACACAGGAAACCAATCAGTGTGAGGACAGATGGGTCTATCTCAGCGGCC	3535
Db	3778	AACGTCATCTGCTCGAGGAACCCAGTCAGTGTGAGGACAGCGGTCTATCCAGCGGCC	3837
QY	3536	TACCCCAACCCGATTCCTATGTCAAAAGGTATGAAAGCTGGAAGCCAGTAGTGGCAGC	3595
Db	3838	TACACCAACCCGATTCCTATGTCAAAAGGTATGAAAGCAGGAAAGCCAGTAGTGGCAGC	3897

QY	3596	CTCAGGAGCAGGAAATCTGACCAAAATTCAGCCTCGAGCTCGAGACTCAGTCTATGAAAAAAT	3655
Db	3898	CCCAGGAGCAGGAAATCTGACCAAATTCGAGCCTCGAGCTCGAGACTCAGTCTATGAAAAAAT	3957
QY	3656	AGAGCTGAAGAAATCTGCAGCCAGCAGCACTGCTCCTCTCTTTGGAGGGGGGAAAGGCTGAGG	3715
Db	3958	AGAGCTGAAGAAATCTGCAGCCAGCAGCACCACTCTCTCGAGGGGGGAAAGGCTGAGG	4017
QY	3716	GCAGTGGCTAAAGGGGTATGTTGTAAAGATGCTACTGCTGCAGTGCAACAACACCTTCCT	3775
Db	4018	GCAGTGGCTAAAGGGGTATGTTGTGCAGATGCTACTGCTGCCGTGAAAGTGAACTTCAT	4077
QY	3776	CTGTGCCAACCCCTTCCCTTGTA-CTACTAATTAAAGTTTTAAATATCTTGTTTATAAAAAT	3834
Db	4078	CTGTTTGTGCCAGTTCCTTACATGTACTAATTTAAAGTTTTAAATATCTTGTTTATAAAAAT	4137
QY	3835	AACCATTTAATAGCCATGCAACCCCCCTCCCATTTTGTGCACTCTGTTTCAATGCAGGGGAA	3894
Db	4138	AACCACTAATAACCAATTGTCT--TTCCCATTTTGTGCACTTGTGTTGATGCTGGGGAA	4195
QY	3895	TAGAATTAATTAGCAGAAATTTCTGTTTGTGTAATGTTCTGTTGAAGATGTTGGTCCAGTT	3954
Db	4196	CA----AAATTAGCAAACTATTGCTTGTGCTTAGAAGCCAGGGCGTGGTTCTAGTTT	4251
QY	3955	CAGTTTACTTCTAGCATGTGCCCCCAATTCAGGTAGTCTACAGAGTTGTGAAGCCCTCAA	4014
Db	4252	CAGTTTGTCTTCTAGCAAGTGGAACCATCAATAGACCCATCTGAG-----CCTG	4300
QY	4015	TATCGTCACCGGAGAGATTTGAGGACCACATTACATATGCTCCCAAAGGCTGGCTCCCAA	4074
Db	4301	TTTTCTCATCAGTTAGATGTGGGACTCAAT--CACACGCTCTTCAAGTCCGGCTCCCAT	4358
QY	4075	TTTTCTTAATTGTAAGCCAACCTTTAATAGACTCAGTTCTGTGAT-TTTTTTTTCCAAAAA	4133
Db	4359	ATTTCTTAATTGCAAGCCAAATTTAATGTACCTTGTTCACAAATAATTTTTTATTAAAAA	4418
QY	4134	AAAAATATTTGAAATAGGACAGAGTTTAAACAGTTGTCAATTTGCACTATCAAGCCATGA	4193
Db	4419	AATCCTATTACAAAATAAGACATCTTTAACTATTGTCAATTTGCCCTCTTCACATCATGA	4478
QY	4194	GTTTGATATATGGTTATAAGAAA-AGAATACTTTTCAGAGCTATCACAGGCTCTCTAAAC	4252
Db	4479	ATTTGCTTTATGTCTGGAAAAAACATCACATAGCTATCACAGGGCTGACCTCTAAAAA	4538
QY	4253	TTTTGGAAAAACAAAAGCCCTAATATGACCTCAGGAAAACAATTTGAACATGAAATAAAA	4312
Db	4539	TTTTGCAAAAACAAAAGGTTCTAAGATGATTTTCAGGAAAAATAATGTGAACATGTAATAAAA	4598
QY	4313	TGGAATGAACGTGTGGAATCTTTAAAAAANAANAANAANAANAANAANAANAANAANA	4355
Db	4599	TGGAATGAATATGGAATCCTAAAAAANAANAANAANAANAANAANAANAANAANAANA	4641
RESULT 4			
AR339574			
LOCUS	AR339574	4892 bp	DNA
DEFINITION	Sequence 1065 from patent US 6569662.	linear	PAT 17-AUG-2003
ACCESSION	AR339574		
VERSION	AR339574.1	GI:33726431	
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 4892)		
TITLE	Tang,Y.T., Zhou,P. and Drmanac,R.T.		
JOURNAL	Nucleic acids and polypeptides		
FEATURES	Patent: US 6569662-A 1065 27-MAY-2003;		
source	Location/Qualifiers		
	1..4892		
	/organism="unknown"		
ORIGIN	/mol_type="genomic DNA"		

Query Match 69.3%; Score 3025.6; DB 6; Length 4892;									
Best Local Similarity 82.5%; Pred. No. 0;									
Matches 3593; Conservative 0; Mismatches 734; Indels 27; Gaps 10;									
QY	1	CCACTGGGTTCTTCAAGGATAAAC	CAGCGGC	-GAAAGAACACACCA	ATTGGTT	-AAGGAG	58		
Db	558	CCACTGGGTTCTTCAAGGATAAAC	TACCTACATAGAGACATAC	CTTTGGTTAAAGGAG	617				
QY	59	TCGACAAACAGGTGGGATGAGAT	CACGAAATCAAGTGGAGAAAG	TTTCATCTAACGGGCA	118				
Db	618	CTGCCGTGAGTGGGATGAGATCT	CGAAACCAAGGTGGTGAAG	TGCATCTGATGGCA	677				
QY	119	TGTCTCCTGCCCAAGTCCCTCC	ATCATCAGCAGTGTGTTAAGG	CCCCCTCAGAAAG	178				
Db	678	TATCTCCTGTCCCAAGCCCTCC	ATCATCGGCAATGCTGGTGA	AAAAAGTCTCTCAGAA	737				
QY	179	TGCAAAAAAGAACAG---	GCCAATCGGAAGGAGGAG	TGTCATGGCTTCCGGA	ACTAT 235				
Db	738	TGCAAAAAAGAACGAAATCA	AATAGGAAGGAGGATGAT	GTCTATGGCCTCAGAA	CTGT 797				
QY	236	CAAAAGGCACCTCAACCATCT	GGAGAAAGTGAGAAAGACT	TAAGAACTGTGGAG	TTT 295				
Db	798	CAAACGACACCTAAAAACAT	CTGGAGAAATGTGAACGAA	AAAAACTAAGAAATCC	TGGAGTT 857				
QY	296	ATCCAAGGAGGACCTCATCC	AGCTCCTGAGTATCATGGA	AGGGAGTTGCAAGCTC	GAGA 355				
Db	858	ATCCAAAGAGACCTCATCC	AACTACTCAGTATATATG	GAAGGGAGTTGCAGS	CCAGAGA 917				
QY	356	AGATGTCTCCACATGCTGAG	GACAGAGAAAAACCAAG	CCCGAGGTTCTGGAGS	CACACTA 415				
Db	918	AGATGTGATCCACATGCTG	AAGACAGAGAAAAACCA	AGCCTGAGGTTCTG	GAGGCTCATTA 977				
QY	416	TGGATCTGCAGAACCTGAG	AAAGTGCTTCGGGTCTC	GCACCGAGATGCCAT	CTCTTGCTCA 475				
Db	978	CGGGTCTGCGGAGCCAG	AGAAAGTGCTGCGGTCT	GCACCGAGATGCCAT	TCTTGCCCA 1037				
QY	476	AGAGAAGTCCATAGGAG	AAGCGTCTATGAGAA	ACCTATCTCAGAGCT	GAGACTGGA 535				
Db	1038	GGAGAAATCCATAGG	AGAGATGTCTATGAG	AAACCGATTTCA	GAGCTGGACAGACTTGA 1097				
QY	536	GGAAGAGCAGAGGAG	ACGTACCGCCGCTAG	AGCAGCTGCTGCT	GGCTGAGAAGTG 595				
Db	1098	GGAAGAACAGAAAG	AACCTACCGCGCAT	GTAGAGCAGCTGT	TGCTGGCCGAGAAGTG 1157				
QY	596	TCACAGGCGCACCGT	PACGAGCTGGAGAAC	GAGACACAGCACAC	TGACTACATGAA 655				
Db	1158	TCATAGGCGCACCG	TATACGATTAGAG	AACGAGAGCATAA	CACACTGACTACATGAA 1217				
QY	656	CAAGAGCGACGACTT	CACCAACCTGCTGG	AGCAGGAGAGGTT	TGAAAAAGCTCCT 715				
Db	1218	CAAGAGCGACGACTT	CACCAACCTGCTGG	AGCAGGAGGAGGTT	TAAAAAAGCTCCT 1277				
QY	716	TGAACAAGAAAAAGCTT	ACCAAGCCCGCAAG	AAAGGAAAAACG	TAAAGCGGCTCAACAA 775				
Db	1278	TGAACAAGAAAAAGCTT	ATCAAGCCCGCAAG	AAAGGAAAAATG	CTAAACGACTCAATAA 1337				
QY	776	ACTTCGAGATGAGCTT	GTAAGCTCAAGTCC	TTCCGCCCTCAT	GTGTTGGTGGACGAGGCA 835				
Db	1338	ACTAAGAGATGAGCTT	GTAAAACTCAATCC	TTTGCACTCAT	GCTGTTGGTGAAGACA 1397				
QY	836	GATGCACATCGAGCA	ACTGGGCTGCAGAG	TCAAGAAAGTCC	AGGACCTCACTCAGAA	GCT 895			
Db	1398	AATGCACATTGAAC	AACTTGGCTGCA	AGCCAGAAAGTAC	AGGATCTTTACTCAGAA	GCT 1457			
QY	896	GAGGAGGAGGAAG	AAAAAACTCAAAG	CGGTCACTTACAA	ATCCAAAGGAAGCC	CCAGAA 955			
Db	1458	GAGGGAAGAAAG	AGAAAGCTCAAAG	CCCTTACTTCCAA	ATCCAAAGAGACAGAC	AGAA 1517			
QY	956	GCTGCTCAAGTTAG	AAGTGGACTTCG	AACACAGGCCTC	GAGGTTTTCC	CAGGAGCACGA 1015			
Db	1518	ATTGCTCAAGTTAG	AAGTGGACTTTG	AACACAGGCCTC	GAGGTTTTCT	CAAGAGCATGA 1577			
QY	1016	AGAGATGAACG	CCAAATTTGG	CGAATCAAGA	ATCTCAACACCG	GCACTTCGACTCAAACT 1075			
Db	1578	AGAGATGAACGCTTAA	ACTGGCTAATCAAG	AGTCTCACAA	TAGGCAACTTAGGCTCAAGCT 1637				
QY	1076	GTTGGCTTATCG	CAAAGGATTGAG	GAGCTGGAAGACCA	ATAAAAAGCCTT	CAGAAAGC 1135			
Db	1638	GTTGGCTTAA	CCCCAAAGAATCG	AGGAGCTAGAGAGAC	CAACAAAAATCT	GCAGAAAGC 1697			
QY	1136	AGAGGAAGAGCTC	CAGGAGCTGAGAG	AGAAAAATTGCC	AAAGGGGAATGT	TGAAAACTCCAG 1195			
Db	1698	AGAGGAAGAACTT	CAAGAAATTAAG	AGATAAAAATTGCC	AAAGGAGAAATGT	TGAAAACTCTAG 1757			
QY	1196	TCTCATGGCGGA	AGTGGAGAGTCT	CGCAAGCGCTGCTT	GAGATGGAGGGCAAG	GATGA 1255			
Db	1758	CCTCATGGCAGA	AGTGGAAAAATCTT	CGAAAGCGTGTGCTT	GAAATGGAAGGT	AAAGATGA 1817			
QY	1256	AGAGATCACGA	AGACCGAGGCCAG	TGTCGCGGAGCTG	AAAGAAAGCTC	CAAGAGGAAGA 1315			
Db	1818	GGAGATCACTA	AAAACTGAATCCC	AGTGTAGGGAATTG	AGGAAGAGCT	TGCAAGAGGAAGA 1877			
QY	1316	ACACCACAGCA	AGGAACCTTAGACT	AGTAGAAGTGG	AGAAAGCTGC	CAGAGAGGATGCTGAGCT 1375			
Db	1878	ACACCATAGT	AAGAGCTCAGACTT	TGAAGTTG	AGAAAGCTAC	CAGAAAGAAATGTCTGAACT 1937			
QY	1376	GGAGAAAGCTG	GAGAAAGCGTT	CAGCCGGAGTAA	GTCCGAAATGC	ACCCAGCTCCATCTGAA 1435			
Db	1938	AGAGAAATTG	AGAAAGCATTTAG	CAAGAGTAAATCT	GAGTGCACCC	AGCTACATTTAAA 1997			
QY	1436	CCTGGAGAA	GGAGAAACCTTA	ACCAAAGACCTG	CTGAACGAGCT	GGAGGTGTCAGAG 1495			
Db	1998	TCTGGAGAA	AGAAAAAGAACTT	AAACCAAGACCTG	CTAAATGAAT	TGGAGGTGTCAGAG 2057			
QY	1496	TCGAGTTAA	AGAACTCGAAATG	CTCCGAGAGTAG	ACTGGAGAGCGG	CAGTTAAGCTCAA 1555			
Db	2058	TCGAGTTAA	AGAAATGGAATG	TTCTGAAA	AGTAGATTG	GAAAAAGCTGAATTAAGCCTAAA 2117			
QY	1556	AGATGACCTT	ACAAAAGCTGA	AGTCCCTTCACT	GTGATGCTG	TGGATGAGAGGAAAAATAT 1615			
Db	2118	AGATGATCTT	ACCAAGTTGA	AGTCATTTAC	CGTGATGCTG	TGTAAGGAAAAATAT 2177			
QY	1616	GATGGAGAAA	ATAAGCAAGAG	AGAGAAAGTGG	ATGAAATAAAAACTTT	TAAGGT 1675			
Db	2178	GATGGAAAA	ATAATAACAAG	AGAGAAAGTGG	ATGAACTCAATAAAAAATTT	TAAGGT 2237			
QY	1676	GGAGCAGG	AAAAAGTCATGG	ATGTGACGGA	AAAGCTAATCGAGG	AAAGCAAGAGCTTTT 1735			
Db	2238	GGAAACAAGG	AAAAAGTTATGG	ATGTAACTG	AAAAAACTAATTTGAAG	AAAGTAGAAAACTTTT 2297			
QY	1736	AAAACTCAA	ATCTGAAATGG	AGAAAAAGGATAC	AGTCTGTGA	ACTGACAAAGGAGAGGATGAGCT 1795			
Db	2298	AAAACTAAA	ATCTGAAATGG	AGAAAAAGTATACA	ACTTGACAAAGAGAAAGAGATGAGTT 2357				
QY	1796	GATGGTAA	AACTGAGGAGCG	GAAGAAAGGTCC	TGTGA	ACTGAGCTGCAGTGTAGACTT 1855			
Db	2358	GATAGGCA	AAATTGAAAA	AGTGAAGAA	AAATCCTCTGA	ATTAAAGCTGCAGTGTGACTT 2417			
QY	1856	ACTAAAGA	AGCGGCTTGATGG	CATAGAGGAGGTAG	AAAAAGGAAAAATAAAC	CCGAGGTAGGTC 1915			
Db	2418	ACTAAAGA	AGAGACTTGATGG	TATAGAGGAAGTGG	AAAGAGAAATAACA	AGAGGAAGGTC 2477			
QY	1916	GTGCAAGG	GGTCTGAGTTCA	CTGCCGGAAGACA	ATAAGATCAGAG	AACTAACCGTTGA 1975			
Db	2478	ACGAAAAG	GGTCTGAGCTCAC	CTGCCGGAAGATAA	TAGATTAAAGGA	AACTAACACTTGA 2537			
QY	1976	AATCGAGAG	ACTGAAGAAA	CGGCTCCAGC	AGTTGGAGGTGG	TGAGGGGACTTGATGAA 2035			
Db	2538	AATTGAGAG	ACTGAAGAAA	CGCTCCCA	CAATTTGGAAGTGG	TGAAAGGGGATTTGATGAA 2597			
QY	2036	GACCGAGG	ACGAATATGACC	AGTTGAGCAGAA	AGTTTCAGAAC	CGGAGCAGGATAAGGCAAA 2095			
Db	2598	GACAGAAG	ATGATGATCAG	CTGGAACAGAA	ATTTAGAACTGAG	CAGGATAAAGGCTAA 2657			
QY	2096	CTTCCTCT	CCCCAGCAGCTC	GAGGAAATCAA	ACCAATGGCCA	AGCAACAAGCCATAGA 2155			

Db 2658 CTTCTCTCTCAACAACTAGAGGAGATCAAGCACCAAAATTGCCAAGAAATAAAGCAATAGA 2717

Qy 2156 GAAAGGGAGGCCGTGAGCCAGGAAGCCGAACCTGCACACAGGTTTCGGCTGGAGGAGGC 2215

Db 2718 GAAGGTTGAGGTTGTGAGCCAGGAAGCTGAACCTGAGACACAGATTTCGGTTGGAAGAAGC 2777

Qy 2216 TAAAGTCTGTATTTACAGGCCGAGGTGCAGGCTCTCAAGGAGAGAATCCACGAGCTGAT 2275

Db 2778 TAAAGTCGAGACTTAAAGCCGAAGTACAAGCTTTAAAGAGAAGATTACGAAATTAAT 2837

Qy 2276 GAACAAAGGAAGACCAAGCTGTCTCAGCTCCAAGTCGACTATTTCGGTCTTTCAGCAAGATT 2335

Db 2838 GAACAAAGGAAGATCAGCTTCTCAGCTCCAGGTAGATTATTCTGTACTTCAACAAAGATT 2897

Qy 2336 TATGGAAGAAGAACTAAGAACAAAGAAATGGGAGGGAGGTCTCTCAATCTGACCAAGGA 2395

Db 2898 TATGGAAGAAGAAATAAGAACAAAAACATGGGCGAGGAGGTTCTCAATCTGACCAAGA 2957

Qy 2396 GCTAGAGCTTTCGAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAAACGGCCGAAGGAT 2455

Db 2958 GTTGAGCTTTCGAAGCGCTACAGCAGAGCTCTTAGGCCAGTGTGAATGGAAGAAGAAAT 3017

Qy 2456 GGTGGACGTGCTGTGGCCCTCCACTGGGTGACAGCCGAGCGGTGTGCGGGGATGCTGC 2515

Db 3018 GGTGGATGTTCTGTGACGTCAACTGGAGTCCAAACTGATGAGTCAAGCGGTGAAGCAGC 3077

Qy 2516 GGAGGAGGAGACCCCGGTGTGTTTCAATTCGCAAAATCTTCCAGGAGGAAATCACATCAT 2575

Db 3078 AGAGGAAGAAACCCAGCTGTATTATCATACGGAATCTCTCCAGGAAGAAATCATATTAT 3137

Qy 2576 GAGTAATCTTCGACAGGTAGGCTGAAGAAACCCATGGAACGGTCTCTCGTCTCGACAG 2635

Db 3138 GAGTAATCTTCGGCAGGTGGGATTGAAGAAACCCGTGGAAGATCTTCTGTTCTAGACAG 3197

Qy 2636 GTATCCCCAGCAGCGAATGAGCTCAACCATGAGGAAGTCTTGGATTCTTGGATGAGAAA 2695

Db 3198 GTATCCTCCAGCAGCAATGAGCTCACTATGAGAAAGTCTTGGATTCCATGGATGAGAAA 3257

Qy 2696 AAGAGAAAAACGGTCTCTTCCACTCCGCAGGAGAAAGGCCCCAGGCCCAACCCAGGGTGCAGG 2755

Db 3258 GAGGAAAAACGGCCCCCTCCATCACTCAGGAGAAAGGGCCCGAAACAAATCCAGTCCAGG 3317

Qy 2756 GCACCCCGGGAGCTGGTCTCTAGCACCAAGCAGGGCCAGCCCCCTACACATCCGTGTGAC 2815

Db 3318 GCACCCAGAGAGGTTAGTCTTTCACCAAAGCAGGGCCAGCCCCCTGCATATTTCGAGTGAC 3377

Qy 2816 ACCAGATCATGAGAACAGCACTGCCACCTGGAGATCAAAAGCCCCACATCTGAAGATT 2875

Db 3378 ACCAGACCAGGAGACAGCACTGCGACTTTGGAGATAACAAAGCCCGACATCTGAAGRAAT 3437

Qy 2876 TTTCTCTAGTACCACCGTCAATCTCTACCTTAGGCAACACAGAAACCAAGAATAACCATTAT 2935

Db 3438 TTTTCTAGTACCCTGTCTCATCTCTA CCTTAGGGAATCAGAAACCAAGAATAACCATTAT 3497

Qy 2936 TCCATCACCCAATGTCTATGTCGAAAAAGCCCCAAAGTCAGATCCTACTCTCGGCCCAGA 2995

Db 3498 TCCATCACAAAACGTTATGCCTCAAAAACAAAAGTGGAGATACTACTCTTGGCCCAGA 3557

Qy 2996 ACGAGCCATGTCCCCCTGTCAAGATTACTACTATTTCAGAGAGAGAAGAGCCCCGGAAGTGG 3055

Db 3558 ACGAGCCATGTCCCCAGTCAAAATTAATACTATTTTCAGAGAGAGAAGACTCCAGAAAAGTGG 3617

Qy 3056 AAGGAGCGCCTTTGCCGACAGGCCTGCATCCCCCATCCAAATCATGACGGTGTCAACATC 3115

Db 3618 AAGAGCGCATTTGACAGACAGGCCCATCCCCCTATTCAGATAATGACGGTGTCTACATC 3677

Qy 3116 TGCAGTCCCACGTGAATCGTGTCTCTCTCTGAATCTCAGGAAGTGCCTATGGGAAGGAC 3175

Db 3678 AGCAGCACAGCTGAGATTGCAGTTTCTCCCGAATCCAGGAAATGCCCATGGGACGGAC 3737

Qy 3176 TATCTCAAAGTCAACCCCGGAAAAACAAACTGTTCCAGCCCCCGGTGCGGAAGTACAACCTC 3235

Db 3738 AATCTCAAAGTCAACCCCGAAAAAACAGACTGTTCCAACTCCAGTACGGAAATACAACCTC 3797

Qy 3236 CAATGCTAATATCATCACCCGGAAGACAATAAAATTCACATTCACCTGGGTCTCAGTT 3295

Db 3798 CAATGCCAATATCATAAACACAGAGGACATAAAATTCACATTCACCTAGGCTCAGTT 3857

Qy 3296 TAAGCGATCTCTGGCCCTGCGCTGAAGGCGTGAGCCCAAGTTATCACCGTCCGGCCTGT 3355

Db 3858 TAAACGGTCCCCTGGACTTCAGGTGAAGGAGTCAGTCCAGTTATTACTGTCCGACCAGT 3917

Qy 3356 CAACGTGACAGCGGAGAAAGGAGGTTTCTACAGGCACAGTCTTCGCTCTCCAGGAACCA 3415

Db 3918 AAACGTGACAGCCGAAAGGAGGTTTCCACCGGCACCTGCTCTCGCTCTCCAGGAATCA 3977

Qy 3416 CCTCTCTTCAAGACCCCGTGTCTAGCAAAAGTGACCACTATAAATACCCCGGTCTAC 3475

Db 3978 CCTCTCTCACGGCTTGGTCAAGCAAAAGTGACGAGCACTATCACCATAAACACCCGTCTAC 4037

Qy 3476 AACGTATCCACACGAGGAACCCAAATCAGTGTCAAGGACAAGATGGGTCTATCAGCGGC 3535

Db 4038 AACGTATCTGCTCGAGGAACCCAGTCAGTGTCAAGACAAGACGGGTCTATCCAGCGGC 4097

Qy 3536 TACCCCCACCCGCTTCTCTATGTCAAAAGGTATGAAAGCTGGAAGCCAGTAGTGGCAGC 3595

Db 4098 TACACCCACCCGCTTCTCTATGTCAAAAGGTATGAAAGCAGGAAGCCAGTAGTGGCAGC 4157

Qy 3596 CTCAGGAGCAGGAATCTGACCAAAATTCAGCCCTCGAGCTGAGACTCAGTCTATGAAAAAT 3655

Db 4158 CCCAGGAGCAGGAATCTGACCAAAATTCGAGCCTCGAGCTGAGACTCAGTCTATGAAAAAT 4217

Qy 3656 AGAGCTGAAGAAATCTGCAGCCAGCAGCACTGCCTCTCTTGGAGGGGGAAGGCTGAGG 3715

Db 4218 AGAGCTGAAGAAATCTGCAGCCAGCAGCACCCTCTCTCGGAGGGGGAAGGCTGAGG 4277

Qy 3716 GCAGTGGCTAAGGGGTATGTTGTAAGGATGCTACTGCTGCAGTGGAACAAACCTTCTCT 3775

Db 4278 GCAGTGGCTAAGGGGTATGTTGTCAGATGCTACTGCTGCCGTGAAAGTGAACCTTCTAT 4337

Qy 3776 CTGTGCCAACCCCTTTCCTTGTA-CTACTAAATTTAAGTTTAAATATCTTGTATATAAAT 3834

Db 4338 CTGTTGTGTCAGTCTCTTACATGTACTAAATTTAAGTTTAAATATGTTGTTTATAAAT 4397

Qy 3835 AACCATTTAATAGCCATGCACCCCTCCCATTTTGTGTCATCTGTTTCAATGCAGGGGAA 3894

Db 4398 AACCAACTAATAACCATTTGTCTT--TCCCATTTTGTGCAATTTGTTTGTGCTGGGAA 4455

Qy 3895 TAGAATTAATTAGCAGAAATTTCTGTTTGTGTAATGTTCTGTTGAAGATGTTGGTCCAGTT 3954

Db 4456 CA---AAATTAGCAAAACTATTGCTTGTCTGCCCTAGAAGCCAGGCGTGGTTCTAGTTC 4511

Qy 3955 CAGTTTACTTCTAGCATGTGGCCCCCATTCAGGTPAGCTCACGAGTTGTGAAGCCCTCAA 4014

Db 4512 CAGTTTGTCTTAGCAAGTGGACCCCATCAATAGACCCATCTGAG-----CCTG 4560

Qy 4015 TATCGTCACCGGAGAGATTTGAGGACCACATTACATATGCTCCCAAAGGCTGGCTCCCAA 4074

Db 4561 TTTCTCTCATCAGTTAGATGTGGGACTCAAT--CACACGCTCTTCAAGTCCGGCTCCCAT 4618

Qy 4075 TTTTCTTAATTGTAAGCCAACTTTAATAGACTCAGTTCTGTGAT-TTTTTTTTCCAAAAA 4133

Db 4619 ATTTCTTAATTGCAAGCCAAATTTAATGTACCTTGTTCACATAAATTTTTTTATTAATAA 4678

Qy 4134 AAAAATATTTTGAATAAGGACAGAGTTTAAAGTGTCAATTTTGCACTATCAAGCCATGA 4193

Db 4679 AATCCTATTACAAAATAAGACATCTTTAACTATTGTCTATTTCCTCTTTTCACATCATGA 4738

Qy 4194 GTTGTATATATGGGTTATAAGAAA-AGAATACTTTTCAGAGCTATCACAGGCTCTCTAAAC 4252

Db 4739 ATTTGCTTTATGTGCTGGAATAAATCATCATAGCTATCACAGGCGCTGGACCTCTAAAA 4798

Qy 4253 TTTTGGAAAAACAAAAAGCCCTTAATATGACCTCAGGAAAAAATTTGAACATGAAATAAAA 4312

Db 4799 TTTTGCAAAAACAAAAAGGTTCTAAGATGATTTTCAGGAAATAATGTGAACATGTAATAAAAA 4858

QY 4313 TGGAAATGAACTGTGGAATCTTAAAAA 4346
|||||
Db 4859 TGGAAATGAAATATGGAATCTTAAAAA 4892
|||||

RESULT 5
AB086011
LOCUS AB086011 4247 bp mRNA linear PRI 03-JUN-2003
DEFINITION Homo sapiens mRNA for L-FILIP, complete cds.
ACCESSION AB086011
VERSION AB086011.1 GI:31338836
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Nagano,T. and Sato,M.
TITLE Human orthologue of L-FILIP
JOURNAL Published Only in Database (2003)
REFERENCE 2 (bases 1 to 4247)
AUTHORS Sato,M. and Nagano,T.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-2002) Makoto Sato, Fukui Medical University,
Department of Anatomy; Shimoaizuki 23, Matsuoka, Fukui 910-1193,
Japan (E-mail:makosato@fmsrsa.fukui-med.ac.jp,
Tel:81-776-61-8305(ex.2205), Fax:81-776-61-8155)
FEATURES
source
1. .4247
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1. .4247
/gene="l-filip"
7. .3648
/gene="l-filip"
/codon_start=1
/product="L-FILIP"
/protein_id="BAC77067.1"
/db_xref="GI:31338837"
/translation="MRSRNOGGESASDGHISCPKPSIIIGNAGEKSLSEDAKKKKSNR
KEDDVMASGIVKRHLKTSGECEKTKKSLELSKEDLIQLLSIMEGLQAREDIVHMLK
TEKTKPEVLEAHYGSAPKVLRLVHRDAI LAQESIGEDVYEKPISELDRLEEKQKE
TYRRMLEQLLAEKCHRRTVYELENEKHKHTDYMNKSDDFTNLLEQERERLKKLLEQE
KAYQARKEKENAKRLNKLRLDELVKLSFALMLVDEQRMHIEQLGQSKVQDLTKQLR
EEEEKKAITSKSKEDRQKLLKLEVDPEHKASRFSQHEHEMNAKLANQESHNRQLRLK
LVGLTQRIELEETNKNLQAEELQELRD KIAKGECSNLSLMAEVLNRKRVLEMEG
KDEEITKTESQCELRKQLQEEHHSKELRLEVEKLQKRMSELEKLEAFSKSSECT
QLHLNLEKXNLT KDLNLELVKSRVKELECSERLEKAELSLXDDLTKLSFTVML
VDERKNMMEKI QKEERKVDGLNKNFVQGVKMDVTEKLI EESKLLKLKSEMEEKVY
NLTRREDELIGLKSEEEKSSELSCSVDLLKRLDGI EEVEREITGRSRKGSSELTCP
EDNKIKELTLEIERLKKRLQOLEVVEGLMKTEDEYDQLEQKFRTEODKANFLSQOLE
EIKHQIAKNKAI EKGEVVSQAEALRHFRLEBAKSRDLKAEVQALKEKIHLMNKEDQ
LSQLOVDYSLVQQRFMEEEKNKNMGQEVNLNLTKELELSKRYSRALRPSVNGRRMVDV
PVTSTGVQTDVSGEAAEETPAVFIRKSFQEEHIMSNLRQVGLKPKPVERSSVLDRY
PPAANELTMRKSWIPWMRKRENGPSITQEKGPRTNSSPGHPGEVVLSPKQGQPLHIRV
TPDHENSTATLEITSPTSEEFFSSTTVITPLGNQKPRITIIIPSNVMPQKQKSGDITLL
GPERAMSPVITITFSREKTPESGRGAFADRPSTPIQIMTVSTSAAPAEIAVSPESQEM
PMGRITLKVTPKQVPTPVKRYNSNANIITEDNKIHIHLGSQFKRSPGTSGEVSP
VITVRPVNVTAEKEVSTGTVLRSRNLSSRPGASKVTSTITITPTVTSSARGTQSVS
GQDSSQRPTPTRIPMKGMKAGKPVVAAPGAGNLTKFEPRAEQTSMKIELKSAANS
TTSLGGGKG"

ORIGIN
Query Match 68.5%; Score 2987.2; DB 9; Length 4247;
Best Local Similarity 82.6%; Pred. No. 0;
Matches 3523; Conservative 0; Mismatches 718; Indels 25; Gaps 8;
QY 69 GTGGGAATGAGATCAGCAATCAAGGTGGAGAAAGTTTCATCTAACGGGCATGTCCTCTGC 128
|||||
Db 1 GTGGGAATGAGATCTCGAAACCAAGGTGGTGAAGTGCAATCTGATGGGCATATCTCCTGT 60
|||||
QY 129 CCCAAGTCCTCCATCATCAGCAGTGATGGTGGTAAGGGCCCCCTCAGAAAGATGCAAAAAAG 188
|||||

Db 61 CCCAAGCCCTCCATCATCGGCAATGCTGGTGAAAAAAGTCTCTCAGAAGATGCAAAAAAG 120
|||||

QY 189 AACAAAG--GCCAATCGGAAGGAGGAGGATGTCTATGSGCTTCCGGAACATATCAAAAGGCAC 245
|||||

Db 121 AAGAAGAAATCAATAGGAAGGAGGATGATGTCTATGGCCTCAGGAACGTGTCAAAACGACAC 180
|||||

QY 246 CTCAAACCATCTGGAGAAAGTGAGAAAAAGACTAAAGAGTCTGTGGAGTTATCCAAAGGAG 305
|||||

Db 181 CTAAAAACATCTGGAGAAATGTGAACGAAAAAACTAAGAAATCCCTGGAGTTATCCAAAGAA 240
|||||

QY 306 GACCTCATCCAGCTCCTGAGTATCATGGAAGGGGAGTTGCAGGCTCGAGAAGATGTCTATC 365
|||||

Db 241 GACCTCATCCAACTACTCAGTATAATGGAAGGGGAGTTGCAGGCCACAGAGAAGATGTGATC 300
|||||

QY 366 CACATGCTGAGGACAGAGAAAAACCAAGCCCGAGGTTCTGGAGGCACACTATGGATCTGCA 425
|||||

Db 301 CACATGCTGAAGACAGAGAAAAACCAAGCCTGAGGTTCTGGAGGCTCATTAACGGGTCTGCG 360
|||||

QY 426 GAACTTGAGAAAGTGTCTCGGGTCTCTGCACCGAGATGCCATCCTTGCTCAAGAGAAGTCC 485
|||||

Db 361 GAGCCAGAGAAAGTGTCTCGGGTCTCTGCACCGAGATGCCATTTCTGCCCCAGGAGAAATCC 420
|||||

QY 486 ATAGGAGAAGACGTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAAAAAGCAG 545
|||||

Db 421 ATAGGAGAAGATGTCTATGAGAAACCGATTTTCAGAGCTGGACAGACTTTGAGGAAAAACAG 480
|||||

QY 546 AAGGAGACGTACCGCCGCATGCTAGAGCAGCTGCTGCTGGCTGAGAAGTGTACAGGCGC 605
|||||

Db 481 AAAGAAACCTACCGGCCCATGCTAGAGCAGCTGTTGCTGGCCGAGAAAGTGTATAGCGCC 540
|||||

QY 606 ACCGTGTACGAGCTGGAGAACGAGAAGCACAAGCAGACACTGACTACATGAACAAGAGCGAC 665
|||||

Db 541 ACCGTATACGAGTTAGAGAACGAGAAGCATAAACACACTGACTACATGAACAAGAGCGAC 600
|||||

QY 666 GACTTCACCAACCTGCTGGAGCAGGAGCGAGAGAGGTTGAAAAAGCTCCTTGAACAAGAA 725
|||||

Db 601 GACTTCACCAACCTGCTGGAGCAGGAGCGGGAGAGGTTAAAAAAGCTCCTTGAACAAGAA 660
|||||

QY 726 AAAGCTTACCAAGCCGCCCAAGAAAAAGGAAAAACGCTAAGCGGCTCAACAAACATTCGAGAT 785
|||||

Db 661 AAGGCTTATCAAGCCGCCCAAGAAAAAGGAAAAATGCTAACGACTCAATAAACTAAGAGAT 720
|||||

QY 786 GAGCTTGTGAAGCTCAAGTCTCTCGCCCTCATGTTGGTGGACGAGAGGCGAGATGCACATC 845
|||||

Db 721 GAGCTTGTAAACTCAAATCCTTTGCACTCATGCTGGTGGATGAAAGACAAATGCACATT 780
|||||

QY 846 GAGCAACTGGGCTGCAGAGTCAAGAAAGTCCAGGACCTCACTCAGAAAGCTGAGGAGGAG 905
|||||

Db 781 GAACAACTTGGCCTGCAAAAGCCAGAAAGTACAGGATCTTACTCAGAAAGCTGAGGGAAGAA 840
|||||

QY 906 GAAGAAAAACTCAAAGCGGTCACTTACAAATCCAAAGGAAGACCGCCAGAAAGCTGCTCAAG 965
|||||

Db 841 GAAGAGAAGCTCAAAGCCATTACTTCCAAATCCAAAGGAAGACAGACAGAAAAATTGCTCAAG 900
|||||

QY 966 TTAGAAGTGGACTTCGAAACACAAGGCCCTCGAGGTTTTCCAGGAGCACGAAGAGATGAAC 1025
|||||

Db 901 TTAGAAGTGGACTTTGAACACAAGGCTTCGAGGTTTCTCAAGAGCATGAAGAGATGAAC 960
|||||

QY 1026 GCCAAATTGGCGGAATCAAGAAATCTCACAACCCGGCAACTTCGACTCAAACTGGTTGGCTTA 1085
|||||

Db 961 GCTAAACTGGCTAATCAAGAGTCTCACAATAGGCAACTTAGACTCAAGCTGGTTGGCTTA 1020
|||||

QY 1086 TCGCAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCCTTCAGAAGGCAGAGGAAGAG 1145
|||||

Db 1021 ACCCAAAGAAATCGAGGAGCTAGAAGAGACCACAAAAAATCTGCAGAAGGCAGAGGAAGAA 1080
|||||

QY 1146 CTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAACCTCCAGTCTCATGGCG 1205
|||||

Db 1081 CTTCAAGAATTAGAGATAAAATTTGCCAAAGGAGAAATGTGGAACCTCTAGCCTCATGGCA 1140
|||||

QY 1206 GAAGTGGAGAGTCTCGCGAAGCGCGTGCTTGTAGATGGAGGCGCAAGGATGAAGAGATCACG 1265
|||||

Db 1141 GAAGTGGAAAAATCTTCGAAAGCGTGTGCTTGAAATGGAAGGTAAAGATGAGGAGATCACT 1200

Qy 1266 AAGACCGAGGCCCAGTGC CGGGAGCTGAAGAAAGCTCAAGAGGAAGAAACACACAGC 1325

Db 1201 AAAACTGAATCCCAGTGTAGGGAATTGAGGAAGAAAGCTGCAAGAGGAAGAACACCATAGT 1260

Qy 1326 AAGGAACCTTAGACTAGAAAGTGGAGAAGCTGCAGAAAGGATGTCTCAGCTGGAGAAGCTG 1385

Db 1261 AAGGAGCTCAGACTTGAAGTTGAGAAGCTACAGAAGAGAATGTCTGAACTAGAGAAATTG 1320

Qy 1386 GAGGAAGCGTTTCAGCCGGAGTAAAGTCCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAG 1445

Db 1321 GAAGAAGCATTTAGCAAGAGTAAATCTGAGTGCACCCAGCTACATTTAAATCTGGAGAA 1380

Qy 1446 GAGAAGAACTTAACCAAGACCTGTCTGAACGAGCTGGAGGTGGTCAAGAGTCAAGTTAAA 1505

Db 1381 GAAAGAACTTAACCAAGACCTGTCTAAATGAATTTGGAGGTGGTCAAGAGTCAAGTTAAA 1440

Qy 1506 GAACTCGAATGCTCCGAGAGTAGACTGGAGAAAGCGAGTTTAAGCCTCAAAGATGACCTT 1565

Db 1441 GAATTTGGAATGTTCTGAAAAGTAGATTGGAAAAGGCTGAATTAAGCCTAAAAGATGATCTT 1500

Qy 1566 ACAAAAGCTGAAGTCCTTCACTGTGATGTGTTGGATGAGAGGAAAAATATGATGGAGAAA 1625

Db 1501 ACCAAGTTGAAGTCATTTACCGTGATGCTGGTTGATGAAGGAAAAATATGATGGAAAAA 1560

Qy 1626 ATAAAGCAAGAGAGAGAAAGTGGATGGGTGGTTGAATAAAAACTTTAAGGTGAGCAGGGA 1685

Db 1561 ATAAAACAAGAGAGAGAAAAGTGGATGGACTCAATAAAAAATTTAAGGTGGAACAAGGA 1620

Qy 1686 AAAGTCATGGATGTGACGGGAAAAGCTAATCGAGGAAAAGCAAGAAAGCTTTTAAAAACTCAA 1745

Db 1621 AAAGTTATGGATGTAACTGAAAAAACTAATTGAAGAAAAGTGAAGAAACTTTTAAAAACTAAAA 1680

Qy 1746 TCTGAAATGGAGGAAAAGGAGTACAGTCTGACAAAGGAGAGGGATGAGCTGATGGGTAAA 1805

Db 1681 TCTGAAATGGAGGAAAAGTATACAACTTGACAAGAGAAAAGAGATGAGTTGATAGGCAAA 1740

Qy 1806 CTGAGGAGCGGAAGAAAGTCTCTGTGAACTGAGCTGCAGTGTAGACTTACTATAAGAG 1865

Db 1741 TTGAAAAGTGAAGAAAGAAAATCCTCTGAATTAAGCTGCAGTGTGACTTACTATAAGAG 1800

Qy 1866 CGGCTTGATGGCATAGAGGAGGTAGAAAAGGGAATTAACCGAGGTAGGTGCTGTAAGGGG 1925

Db 1801 AGACTTGATGGTATAGAGGAGTGGAAAGAGAAATAACAAGAGGAAGGTCAACGAAAAGG 1860

Qy 1926 TCTGAGTTCACCTGCCCCGGAAGACATAACATCAGAGAACTAACGCTTGAAATCGAGAGA 1985

Db 1861 TCTGAGCTCACCTGCCCCGGAAGATAATAAGATTAAAGAACTAACCTTGAAATTGAGAGA 1920

Qy 1986 CTGAAGAAACGGCTCCAGCAGTTGGAGTGGTGGAGGGGACTTTGATGAAGACCGAGGAC 2045

Db 1921 CTGAAGAAACGCTCTCCAACAATTTGGAAGTGGTCTGAAGGGGATTTGATGAAGACAGAAGAT 1980

Qy 2046 GAATATGACCAGTTGGAGCAGAAAGTTTCAGAAACCGAGCAGGATAAGGCAAACTTCCTCTCC 2105

Db 1981 GAGTATGATCAGCTGGAACAGAAATTTAGAACTGAGCAGGATAAGGCTAACTTCTCTCT 2040

Qy 2106 CAGCAGCTCAGGAAATCAAAACACCAAAATGGCCAAAGCACAAAGCCATAGAGAAAGGGAG 2165

Db 2041 CAACAACTAGAGGAGATCAAGCACCAAAATTGCCAAGAAATAAAGCAATAGAGAAGGTGAG 2100

Qy 2166 GCCGTGAGCCAGGAAGCCGAACCTGCGACACAGGTTTCGGCTGGAGAGGCTAAAGTCGT 2225

Db 2101 GTTGTGAGCCAGGAAGCTGAAGCTAGACACAGATTTCCGGTTTGGAAAGCTTAAAGTCGA 2160

Qy 2226 GATTACAGGCCGAGGTGAGGCTCTCAAGGAGAGAAGATCCACGAGCTGATGAACAAGGAA 2285

Db 2161 GACTTAAAAGCCGAAGTACAAGCTCTTAAAGAGAGAAGATTCACGAATTAATGAACAAGAA 2220

Qy 2286 GACCAGCTGTCTCAGCTCCAAGTCGACTATTCCGTCCTTCAGCAAGATTTATGGAAGAA 2345

Db 2221 GATCAGCTTCTCAGCTCCAGGTAGATTATTCTGTACTTCAACAAGATTTATGGAAGAA 2280

Qy 2346 GAAACTAAGAAACAAGAAACATGGGAGGGAGGTCTCTCAATCTGACCAAGGAGCTAGAGCTT 2405

Db 2281 GAAAAATAAGAAACAACAAACATGGGCGAGGAGGTCTCTCAATCTGACCAAGAGTTGGAGCTT 2340

Qy 2406 TCCAAAGCGTTACAGCCGAGCTCTCAGGCCGAGTGGGAAACGSCCGAAGGATGTTGGACGTG 2465

Db 2341 TCCAAAGCGTTACAGCAGAGCTCTTAGGCCAGTGTGAATGGAAGAAGAAATCGTGGATGTT 2400

Qy 2466 CCTGTGGCCTCCACTGGGTGCAGACCGAGCGGTGTCCGGGGATGCTCGGAGGAGGAG 2525

Db 2401 CCTGTGACGTCAACTGGAGTCCAAACTGATGCAGTCAGCCGTGAAGCAGAGGAAGAA 2460

Qy 2526 ACCCCGGCTGTGTTCAATTCGCAAAATCCTTCCAGGAGGAAAAATCACATCATGAGTAATCTT 2585

Db 2461 ACGCCAGCTGATTTCAATACGGAAATCCTTCCAGGAAGAAAAATCATATTATGAGTAATCTT 2520

Qy 2586 CGACAGGTAGGCCTGAAGAAAACCCATGGAACGGTCTCTCGGTCTCTCGACAGGTATCCCCCA 2645

Db 2521 CGGCAGGTGGGATTGAAGAAAACCCGTGGAAGATCTTCTGTCTAGACAGGTATCCTCCA 2580

Qy 2646 GCAGCGAATGAGCTCACCATGAGGAAGTCTTTGGATTCTTGGATGAGAAAAAGAAAAAC 2705

Db 2581 GCAGCAATGAGCTCACTATGAGAAAGTCTTGGATTCCATGGATGAGAAAAGGGAANAAC 2640

Qy 2706 GGTCTTCCACTCCGCAAGGAGAAAGGCCCGAGGCCAAACAGGGTGCAGGGCACCCCGGG 2765

Db 2641 GGCCCTCCATCACTCAGGAGAAAGGGCCCCGAAACAAATTCAGTCCAGGGCACCCAGGA 2700

Qy 2766 GAGCTGCTCTAGCACCAAGCAGGGCCAGCCCCCTACACATCCGTGTGACACCAGATCAT 2825

Db 2701 GAGGTAGTCTTTTACCAAAAGCAGGGCCAGCCCCCTGCATATTTCGAGTGACACCAGACCAC 2760

Qy 2826 GAGAACAGCACTGCCACCTGGAGATCAAAAGCCCCCACATCTGAAGAGTTTTCTCTAGT 2885

Db 2761 GAGAACAGCACTGCGACTTTGGAGATAAACAAAGCCCGACATCTGAAGAAATTTTTTCTAGT 2820

Qy 2886 ACCACCGTCATTCCTTACCTTAGGCAACCCAGAAACCAAGAATAACCATTTATTCATCACCC 2945

Db 2821 ACCACTGTCTATTCCTTACCTTAGGGAAATCAGAAACCAAGAATAACCATTTATTCATCACCA 2880

Qy 2946 AATGTCATGTCGCAAAAAGCCCCAAAAGTGCAGATCCTACTCTCGGCCAGAACGAGCCATG 3005

Db 2881 AACGTTATGCCCTCAAAAACAAAAAAGTGGAGATACTACTCTTGGCCAGAACGAGCCATG 2940

Qy 3006 TCCCTGTGCAGATTACTACTATTTCAGAGAGAGAGCCCCGGAAGGTGGAAGGAGCGCC 3065

Db 2941 TCCCCAGTCACAATTACTACATTTCCAGAGAGAGAAGACTCCAGAAAAGTGGAAAGGCGCA 3000

Qy 3066 TTTGCCGACAGGCCTGCATCCCCCATCCAAATCATGACGGTGTCAACATCTGCAGTCCC 3125

Db 3001 TTTGCAGACAGGCCCCACATCCCCCTATTTCAGATAATGACGGTGTCTACATCAGCAGCACCA 3060

Qy 3126 ACTGAAATCGTGTCTCTCTGAAATCTCAGGAAGTGCCTATGGGAAGGACTATCCTCAAA 3185

Db 3061 GCTGAGATTGCAGTTTCTCCCGAATCCAGGAATGCCCATGGGACGGACAATCCTCAAA 3120

Qy 3186 GTCACCCCGGAAAAACAAACTGTTCCAGCCCCCGTCCGGAAGTACAACTCCAATGCTAAT 3245

Db 3121 GTCACCCCGGAAAAACAGACTGTTCCAACTCCAGTACGGAATACAACTCCAAATGCCAAT 3180

Qy 3246 ATCATCACCGGAAGACAATAAAAAATTCACATTCACTTACCTGGGTTCTCAGTTTAAAGCGATCT 3305

Db 3181 ATCATAAACAGAGGACAATAAAATTCACATTCACTTAGGTTCTCAGTTTAAACCGTCC 3240

Qy 3306 CCTGGCCTGCCGTGAAGCGTGAAGCCAGTCCAGTTTATCACCCGTCCGGCTGTCAACCGTGACA 3365

Db 3241 CCTGGACTTCAGGTGAAGGAGTCAGTCCAGTTATTACTGTCCGACCCAGTAAACCGTGACA 3300

Qy 3366 GCGGAGAAAGGAGTTTCTACAGGCACAGTCTTCGCTCTCCAGGAAACCACTCTCTTCA 3425

Db 3301 GCCGAAAAGGAGGTTTCCACCGGCACGTGTCTTCGCTCTCCAGGAATCACCTCTCTCTCA 3360

QY	3426	AGACCCGGTGCTAGCAAAAGTCACCAGCACTATAACTATAACCCCGGTCACAACGTCATCC	3485
Db	3361	CGGCCTGGTCAAGCAAAGTCAGGAGCACTATCACCATAACACCGGTCACAACGTCATCT	3420
QY	3486	ACACGAGGAACCCCAATCAGTGTCTAGGACAAGATGGGTCTATCTCAGCGGCCTACCCCCACC	3545
Db	3421	GCTCGAGGAACCCAGTCAGTGTCTAGGACAAGACGGGTCTATCCAGCGGCCTACACCCACC	3480
QY	3546	CGCATTCCTATGTCAAAAGGTATGAAAGCTGGAAGCCAGTAGTGGCAGCCTCAGAGCA	3605
Db	3481	CGCATTCCTATGTCAAAAGGTATGAAAGCAGGAAAGCCAGTAGTGGCAGCCCCAGAGCA	3540
QY	3606	GGAAATCTGACCAAAATTCAGCCTCGAGCTGAGACTCAGTCTATGAAAATAGAGCTGAAG	3665
Db	3541	GGAAATCTGACCAAAATTCAGCCTCGAGCTGAGACTCAGTCTATGAAAATAGAGCTGAAG	3600
QY	3666	AAATCTGCAGCCAGCAGCACTGCCTCTCTTGGAGGGGGAAGGCTGAGGGCAGTGGCTA	3725
Db	3601	AAATCTGCAGCCAGCAGCACCCTCTCTCGGAGGGGGAAGGCTGAGGGCAGTGGCTA	3660
QY	3726	AGGGGGTATGTTGTAAAGTACTGCTGCTGCACTGGGAAACAAACCTTCCTCTGTGCCAAC	3785
Db	3661	AGGGGGTATGTTGTGCAGATGCTACTGCTGCCGTGGAAGTGAACCTTCATCTGTTGTGC	3720
QY	3786	CCTTTCCTTGAC-TACTAAATTAAGTTTAAATATCTGTTTATAAAATAACCAATTA	3844
Db	3721	CAGTTCTTTACATGTAATAATTAAGTTTAAATATGTTTATAAAATAACCAACTAA	3780
QY	3845	TAGCCATGCAACCCCTCCCATTTTGTGCATCTGTTCAATGTCAGGGGAATAGAAATTA	3904
Db	3781	TAACCATTTGTCTT--TCCATTTTGTGCATTTGTTGATGCTGGGAACA----AAAT	3834
QY	3905	TAGCAGAATTTCTGTTTGTGTAATGTTCTGTTGAGATGTTGGTCCAGTTCAGTTTACT	3964
Db	3835	TAGCAAAACTATTGCTTGCTGCCTAGAGCCAGGCGGTGTTTCTAGTTCAGTTTGCT	3894
QY	3965	TCTAGCATGTGGCCCATTCAGGTAGCTCACGAGTGTGAAGCCCTCAATATCGTCACC	4024
Db	3895	TCTAGCAAGTGGACCCATCAATAGACCCATCTGAG-----CCTGTTTCTCATC	3943
QY	4025	GGAGAGATTGGAGGACCACATTACATATGCTCCCAAGGCTGGCTCCCAATTTTCCTAAT	4084
Db	3944	AGTTAGATGTGGGACTCAAT--CACACGCTCTTCAAGTCCGCTCCCATATTTCTTAAT	4001
QY	4085	TGTAAGCCAACTTTAATAGACTCAGTTCCTGTGAT-TTTTTCCTCAAAAAAATATTT	4143
Db	4002	TGCAAGCCAAATTTATGTACCTTGTTCCACAAATAATTTTATTAAAAAATCCTATTA	4061
QY	4144	TGAAATAGGACAGAGTTTAAACAGTTGTCTATTTTGCACTATCAAGCCATGATTTGATATA	4203
Db	4062	CAAAATAAGACATACTTTAACTATTGTCTATTTGCTCTTTTCACATCATGATTTGCTTTA	4121
QY	4204	TGGGTTATAAGAAA-AGAATACTTTTCAGAGCTATCACAGGGTCTCTAAACCTTTGGAAAA	4262
Db	4122	TGTGCTGGAAAAACATCACATAGCTATCACAGGGCTGGACCTCTAAAAATTTGCAAAA	4181
QY	4263	ACAAAAGCCCTAATATGACCTCAGGAAACAAATTTGAACATGAAATAAAAATGGAATGAA	4322
Db	4182	ACAAAAGGTTCTAAGATGATTTTCAGGAAATAATGTGAACATGTAAATAAAATGGAATGAA	4241
QY	4323	CTGTGG	4328
Db	4242	ATATGG	4247
RESULT	6		
AC130850			
LOCUS	AC130850	227874 bp	DNA linear HTG 10-MAY-2003
DEFINITION	Rattus norvegicus clone CH230-83M15, WORKING DRAFT SEQUENCE.		
ACCESSION	AC130850		
VERSION	AC130850.3	GI:305222010	
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	Rattus norvegicus (Norway rat)		

ORGANISM	Rattus norvegicus
REFERENCE	1 (bases 1 to 227874)
AUTHORS	Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,P., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Georegeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemeleh,O., Okwuonu,G., Olarpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puozo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 227874)
AUTHORS	Rat Genome Sequencing Consortium.
TITLE	Direct Submission
JOURNAL	Submitted (15-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 227874)
AUTHORS	Rat Genome Sequencing Consortium.
TITLE	Direct Submission
JOURNAL	Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On May 10, 2003 this sequence version replaced gi:23266240. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas

Db	197743	ATAAACCGAGGTAGGTCGTGCAAGGGGTCTGAGTTACCTGCCCCGAAGACAATAAGATC	197802
QY	1959	AGAGAACTAACGCTTGAATTCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTGGTG	2018
Db	197803	AGAGAACTAACGCTTGAATTCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTGGTG	197862
QY	2019	GAGGGGGACTTGATGAAGACCGGAGGACGAATATGACAGTTGGAGCAGAAGTTCAGAAACC	2078
Db	197863	GAGGGGGACTTGATGAAGACCGGAGGACGAATATGACAGTTGGAGCAGAAGTTCAGAAACC	197922
QY	2079	GAGCAGGATAAGGCAACTTCCTCTCCCAGCAGCTCGAGGAAATCAAAACACCAATGGCC	2138
Db	197923	GAGCAGGATAAGGCAACTTCCTCTCCCAGCAGCTCGAGGAAATCAAAACACCAATGGCC	197982
QY	2139	AAGCACAAAGCCATAGAGAAAGGGAGGCCGTGAGCCAGGAAGCCGAACCTCGGACACAGG	2198
Db	197983	AAGCACAAAGCCATAGAGAAAGGGAGGCCGTGAGCCAGGAAGCCGAACCTCGGACACAGG	198042
QY	2199	TTTCGGCTGGAGGAGGCTAAAAAGTCGTGATTTACAGSCCGAGGTGCAGGCTCTCAAGGAG	2258
Db	198043	TTTCGGCTGGAGGAGGCTAAAAAGTCGTGATTTACAGSCCGAGGTGCAGGCTCTCAAGGAG	198102
QY	2259	AAGATCCACGAGCTGATGAACAAGGAAGACCAAGCTGTCTCAGCTCCAAGTCGACTATTGG	2318
Db	198103	AAGATCCACGAGCTGATGAACAAGGAAGACCAAGCTGTCTCAGCTCCAAGTCGACTATTGG	198162
QY	2319	GTCCCTTCAGCAAAAGATTATGGAAGAAGAAACTAAGAAACAAGAACATGGGGAGGAGGTC	2378
Db	198163	GTCCCTTCAGCAAAAGATTATGGAAGAAGAAACTAAGAAACAAGAACATGGGGAGGAGGTC	198222
QY	2379	CTCAATCTGACCAAGGAGCTAGAGCTTTTCCAAGCGCTACAGCCGAGCTCTCAGSCCGAGT	2438
Db	198223	CTCAATCTGACCAAGGAGCTAGAGCTTTTCCAAGCGCTACAGCCGAGCTCTCAGSCCGAGT	198282
QY	2439	GGGAACGGCCGAAGGATGGTGACGTGCCTGTGGCCCTCCACTGGGTGCAGACCGAGCGG	2498
Db	198283	GGGAACGGCCGAAGGATGGTGACGTGCCTGTGGCCCTCCACTGGGTGCAGACCGAGCGG	198342
QY	2499	GTGTGCGGGGATGCTGCGGAGGAGAGACCCCGGCTGTGTTTATTTCGAAATCCTTCCAG	2558
Db	198343	GTGTGCGGGGATGCTGCGGAGGAGAGACCCCGGCTGTGTTTATTTCGAAATCCTTCCAG	198402
QY	2559	GAGGAAATCACATCATGAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCATGGAACGG	2618
Db	198403	GAGGAAATCACATCATGAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCATGGAACGG	198462
QY	2619	TCCTCGGTCTCGACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTGG	2678
Db	198463	TCCTCGGTCTCGACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTGG	198522
QY	2679	ATTCTTGGATGAGAAAAAGAGAAAAACGGTCTTCCACTCCGAGGAGAAAGGCCCCAGG	2738
Db	198523	ATTCTTGGATGAGAAAAAGAGAAAAACGGTCTTCCACTCCGAGGAGAAAGGCCCCAGG	198582
QY	2739	CCAAACCGGGTGCAGGCACCCCGGGAGCTGGTCTTAGCACCAAAAGCAGGGCCAGCCCC	2798
Db	198583	CCAAACCGGGTGCAGGCACCCCGGGAGCTGGTCTTAGCACCAAAAGCAGGGCCAGCCCC	198642
QY	2799	CTACACATCCGTGTGACACCAAGATCATGAGAACAGCACTGCCACCCTGGAGATCACAAAGC	2858
Db	198643	CTACACATCCGTGTGACACCAAGATCATGAGAACAGCACTGCCACCCTGGAGATCACAAAGC	198702
QY	2859	CCCACATCTGAAGAGTTTCTCTAGTACCACCGTCATTCCTACCTTAGGCAACCAAGAA	2918
Db	198703	CCCACATCTGAAGAGTTTCTCTAGTACCACCGTCATTCCTACCTTAGGCAACCAAGAA	198762
QY	2919	CCAAGAATAACCATTTATCCATCACCCCAATGTCTATGTCGCAAAAGCCCAAAAGTCAGAT	2978
Db	198763	CCAAGAATAACCATTTATCCATCACCCCAATGTCTATGTCGCAAAAGCCCAAAAGTCAGAT	198822
QY	2979	CCTACTCTCGGCCCAAGACGAGCCATGTCCCTGTACCGATTACTACTATTTCAGAGAG	3038
Db	198823	CCTACTCTCGGCCCAAGACGAGCCATGTCCCTGTACCGATTACTACTATTTCAGAGAG	198882

QY	3039	AAGAGCCCGGAAGGTGGAAGGAGCGCCTTTGCCGACAGGCCTGCATCCCCCATCAAATC	3098
Db	198883	AAGAGCCCGGAAGGTGGAAGGAGCGCCTTTGCCGACAGGCCTGCATCCCCCATCAAATC	198942
QY	3099	ATGACGGTGTCAACATCTGCAGTCTCCCACTCCCACTGAATCGTGTCTCTCCTGAATCTCAGGAA	3158
Db	198943	ATGACGGTGTCAACATCTGCAGTCTCCCACTCCCACTGAATCGTGTCTCTCCTGAATCTCAGGAA	199002
QY	3159	GTGCCTATGGGAAGGACTATCCTCAAAGTCAACCCCGGAAAAACAACACTGTTCCAGCCCCC	3218
Db	199003	GTGCCTATGGGAAGGACTATCCTCAAAGTCAACCCCGGAAAAACAACACTGTTCCAGCCCCC	199062
QY	3219	GTGCGGAAGTACAACCTCCAATGCTAATATCATCACCGGAAGACAATAAAATTCACATT	3278
Db	199063	GTGCGGAAGTACAACCTCCAATGCTAATATCATCACCGGAAGACAATAAAATTCACATT	199122
QY	3279	CACCTGGTTCTCAGTTTAAGCGATCTCCTGGGCTGCCGTGAAGGCGTGAGCCAGTT	3338
Db	199123	CACCTGGTTCTCAGTTTAAGCGATCTCCTGGGCTGCCGTGAAGGCGTGAGCCAGTT	199182
QY	3339	ATCACCGTCCGGCCTGTCAACGTGACAGCGGAGAAGGAGTTTCTACAGGCACAGTCCTT	3398
Db	199183	ATCACCGTCCGGCCTGTCAACGTGACAGCGGAGAAGGAGTTTCTACAGGCACAGTCCTT	199242
QY	3399	CGCTCTCCCAGGAACCACTCTCTTCAAGACCCCGTCTAGCAAAGTGACCAAGCACTATA	3458
Db	199243	CGCTCTCCCAGGAACCACTCTCTTCAAGACCCCGTCTAGCAAAGTGACCAAGCACTATA	199302
QY	3459	ACTATAACCCCGTCAACAGTCAACAGTCAACAGGAAACCAATCAGTG	3506
Db	199303	ACTATAACCCCGTCAACAGTCAACAGTCAACAGGAAACCAATCAGTG	199350

RESULT 7

AC108529/c

LOCUS

DEFINITION Rattus norvegicus clone CH230-6K7, WORKING DRAFT SEQUENCE, 3 unordered pieces.

ACCESSION AC108529

VERSION AC108529.6 GI:30467897

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 238626)

AUTHORS

Muzny,D., Marie., Metzker,M., Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,

Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smales, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., Von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Query Match 64.3%; Score 2804.8; DB 2; Length 238626;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2806; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 699 AGGTTGAAAAAGCTCCTTGAACAAGAAAAAGCTTACCAAGCCCGCAAGAAAAAGGAAAAAC 758
Db 204156 AGGTTGAAAAAGCTCCTTGAACAAGAAAAAGCTTACCAAGCCCGCAAGAAAAAGGAAAAAC 204097
QY 759 GCTAAGCGGCTCAACAAACTTCGAGATGAGCTTGTGAAGCTCAAGTCTCGCCCTCATG 818
Db 204096 GCTAAGCGGCTCAACAAACTTCGAGATGAGCTTGTGAAGCTCAAGTCTCGCCCTCATG 204037
QY 819 TTGGTGGACGAGAGGCAGATGCATCGAGCAACTGGGCCCTGCAGAGTCAGAAAGTCCAG 878
Db 204036 TTGGTGGACGAGAGGCAGATGCATCGAGCAACTGGGCCCTGCAGAGTCAGAAAGTCCAG 203977
QY 879 GACCTCACTCAGAAAGCTGAGGGAGGAGGAAGAAAACTCAAAGCGGTCACTTACAAATCC 938
Db 203976 GACCTCACTCAGAAAGCTGAGGGAGGAGGAAGAAAACTCAAAGCGGTCACTTACAAATCC 203917
QY 939 AAGGAAGACCGCCAGAAAGCTGCTCAAGTTAGAGTGGACTTCGAACACAAGGCGCTCGAGG 998
Db 203916 AAGGAAGACCGCCAGAAAGCTGCTCAAGTTAGAGTGGACTTCGAACACAAGGCGCTCGAGG 203857
QY 999 TTTTCCCAGGAGCAGAGAGATCAACGCCAAATTGGCGAATCAAGAAATCTCACACCGG 1058
Db 203856 TTTTCCCAGGAGCAGAGAGATCAACGCCAAATTGGCGAATCAAGAAATCTCACACCGG 203797
QY 1059 CAACTTCGACTCAAACTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGAAGAGACCAAT 1118
Db 203796 CAACTTCGACTCAAACTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGAAGAGACCAAT 203737
QY 1119 AAAAGCCTTCAGAAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAATTTGCCAAAGGG 1178
Db 203736 AAAAGCCTTCAGAAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAATTTGCCAAAGGG 203677
QY 1179 GAATGTGAAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTCTTGAG 1238
Db 203676 GAATGTGAAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTCTTGAG 203617

ORIGIN

FEATURES

source

misc_feature

misc_feature

misc_feature

misc_feature

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 236254: contig of 236254 bp in length
* 236255 236354: gap of unknown length
* 236355 237360: contig of 1006 bp in length
* 237361 237460: gap of unknown length
* 237461 238626: contig of 1166 bp in length.
Location/Qualifiers
1. .238626
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-6K7"
1. .1401
/note="wgs end extension"
clone_end:Sp6
1469. .2695
/note="wgs end extension"
clone_end:Sp6
3858. .4635
/note="clone boundary"
clone_end:Sp6
site:EcoRI
end_sequence: BH361027"
129792. .235366
/note="clone boundary"
clone_end:T7
site:EcoRI
end_sequence: BH361024"

QY	1239	ATGAGGGCAAGGATGAAGAGATCACGAAGACCGAGGCCCCAGTGCCGGGAGCTGAAGAAG	1298
Db	203616	ATGGAGGGCAAGGATGAAGAGATCACGAAGACTGAGGCCCCAGTGCCGGGAGCTGAAGAAG	203557
QY	1299	AAGCTCCAAGAGGAAGAACACACAGCAAGGAACCTTAGACTAGAAAGTGGAGAGCTGCAG	1358
Db	203556	AAGCTCCAAGAGGAAGAACACACAGCAAGGAACCTTAGACTAGAAAGTGGAGAGCTGCAG	203497
QY	1359	AAGAGGATGCTGAGCTGGAGAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGC	1418
Db	203496	AAGAGGATGCTGAGCTGGAGAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGC	203437
QY	1419	ACCCAGCTCCATCTGAACCTGGAGAAGGAGGAAGAACCTTAACCAAGACCTGCTGAACGAG	1478
Db	203436	ACCCAGCTCCATCTGAACCTGGAGAAGGAGGAAGAACCTTAACCAAGACCTGCTGAACGAG	203377
QY	1479	CTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAAATGCTCCGAGAGTAGACTGGAGAAG	1538
Db	203376	CTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAAATGCTCCGAGAGTAGACTGGAGAAG	203317
QY	1539	GCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAGTCTTCACTGTGATGCTGGTG	1598
Db	203316	GCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAGTCTTCACTGTGATGCTGGTG	203257
QY	1599	GATGAGAGGAAAAATATGATGGAGAAAAATAAGCAAGAAGAGAGGAAAGTGGATGGTTG	1658
Db	203256	GATGAGAGGAAAAATATGATGGAGAAAAATAAGCAAGAAGAGAGGAAAGTGGATGGTTG	203197
QY	1659	AATAAAAACTTTAAGGTGGAGCAGGGAAAAAGTCAATGATGTGACGGAAAAAGCTAATCGAG	1718
Db	203196	AATAAAAACTTTAAGGTGGAGCAGGGAAAAAGTCAATGATGTGACGGAAAAAGCTAATCGAG	203137
QY	1719	GAAGCAAGAAAGCTTTTAAAACTCAAATCTGAATCTGAATCTGAATCTGAATCTGAATCTGA	1778
Db	203136	GAAGCAAGAAAGCTTTTAAAACTCAAATCTGAATCTGAATCTGAATCTGAATCTGAATCTGA	203077
QY	1779	AAGGAGAGGGATGAGCTGATGGGTAAA CTGAGGAGCCAAAGAAAGAGTCTGTGAAC TG	1838
Db	203076	AAGGAGAGGGATGAGCTGATGGGTAAA CTGAGGAGCCAAAGAAAGAGTCTGTGAAC TG	203017
QY	1839	AGCTGCAGTGTAGACTTACTAAAGAAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGGGAA	1898
Db	203016	AGCTGCAGTGTAGACTTACTAAAGAAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGGGAA	202957
QY	1899	ATAAACCGAGGTAGTCTGTCAAGGGGTCTGAGTTCACTCCCGGAAGACATAAGATC	1958
Db	202956	ATAAACCGAGGTAGTCTGTCAAGGGGTCTGAGTTCACTCCCGGAAGACATAAGATC	202897
QY	1959	AGAGAACTAACGCTTGAAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTGGTG	2018
Db	202896	AGAGAACTAACGCTTGAAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTGGTG	202837
QY	2019	GAGGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAGCAGAAAGTTCAGAAACC	2078
Db	202836	GAGGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAGCAGAAAGTTCAGAAACC	202777
QY	2079	GAGCAGGATAAGGCAAACTTCTCTCCAGCAGCTCGAGGAAATCAAACACCAAATGGCC	2138
Db	202776	GAGCAGGATAAGGCAAACTTCTCTCCAGCAGCTCGAGGAAATCAAACACCAAATGGCC	202717
QY	2139	AAGCACAAGCCATAGAGAAAGGGGAGGCGGTGAGCCAGGAAGCCGAACCTGCGACACAGG	2198
Db	202716	AAGCACAAGCCATAGAGAAAGGGGAGGCGGTGAGCCAGGAAGCCGAACCTGCGACACAGG	202657
QY	2199	TTTCGGCTGGAGGAGGCTAAAAGTCGTGATTTTACAGCCCGAGGTGCAGGCTCTCAAGGAG	2258
Db	202656	TTTCGGCTGGAGGAGGCTAAAAGTCGTGATTTTACAGCCCGAGGTGCAGGCTCTCAAGGAG	202597
QY	2259	AAGATCCACGAGCTGATGAACAAGGAAGACCAAGTGTCTCAGCTCCAAGTGCATATTTCG	2318
Db	202596	AAGATCCACGAGCTGATGAACAAGGAAGACCAAGTGTCTCAGCTCCAAGTGCATATTTCG	202537
QY	2319	GTCTCTCCAGCAAGATTTTATGGAAGAGAAACTTAAGAA CAAGAACATGGGGAGGAGGTC	2378

Db	202536	GTCTCTCAGCAAAAGATTTATGGAAGAAGAAACTAAGAAACAAGAACATGGGGAGGAGGTC	202477
QY	2379	CTCAATCTGACCAGGAGCTAGAGCTTTCCAAAGCGCTACAGCCGAGCTCTCAGCCGAGT	2438
Db	202476	CTCAATCTGACCAGGAGCTAGAGCTTTCCAAAGCGCTACAGCCGAGCTCTCAGCCGAGT	202417
QY	2439	GGGAACGGCCGAAGGATGGTGGACGTGCCTGTGGCCTCCACTGGGGTGCAGACCGAGGCG	2498
Db	202416	GGGAACGGCCGAAGGATGGTGGACGTGCCTGTGGCCTCCACTGGGGTGCAGACCGAGGCG	202357
QY	2499	GTGTGCGGGGATGCTGCGGAGGAGGAGACCCCGGCTGTGTTTCTTCGAAAATCTTCCAG	2558
Db	202356	GTGTGCGGGGATGCTGCGGAGGAGGAGACCCCGGCTGTGTTTCTTCGAAAATCTTCCAG	202297
QY	2559	GAGGAAAAATCACATCATGAGTAATCTTCGACAGGTAGGCCTGAAGAAAACTTGGAAACGG	2618
Db	202296	GAGGAAAAATCACATCATGAGTAATCTTCGACAGGTAGGCCTGAAGAAAACTTGGAAACGG	202237
QY	2619	TCCTCGGTCTCGACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTGG	2678
Db	202236	TCCTCGGTCTCGACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTGG	202177
QY	2679	ATTCTTTGGATGAGAAAAAGAGAAAAACCGTCTTCCACTCCGCAGGAGAAAAGGCCCCAGG	2738
Db	202176	ATTCTTTGGATGAGAAAAAGAGAAAAACCGTCTTCCACTCCGCAGGAGAAAAGGCCCCAGG	202117
QY	2739	CCAAACCCAGGTGCGAGGGCACCCCGGGGAGTGGTCTTAGCACCAAAGCAGGGCCAGCCC	2798
Db	202116	CCAAACCCAGGTGCGAGGGCACCCCGGGGAGTGGTCTTAGCACCAAAGCAGGGCCAGCCC	202057
QY	2799	CTACACATCCGTGTGACACCAGATCATGAGAAACAGCACCTGCCACCTGGAGATCACAAGC	2858
Db	202056	CTACACATCCGTGTGACACCAGATCATGAGAAACAGCACCTGCCACCTGGAGATCACAAGC	201997
QY	2859	CCCACATCTGAAGAGTTTCTCTAGTACCAACCGTCATTCCTACCTTAGGCAACCCAGAAA	2918
Db	201996	CCCACATCTGAAGAGTTTCTCTAGTACCAACCGTCATTCCTACCTTAGGCAACCCAGAAA	201937
QY	2919	CCAAAGAAATACCATTATTCCATCACCCCAATGTTCATGTCGCAAAAAGCCCCAAAAGTGCAGAT	2978
Db	201936	CCAAAGAAATACCATTATTCCATCACCCCAATGTTCATGTCGCAAAAAGCCCCAAAAGTGCAGAT	201877
QY	2979	CCTACTCTCGGCCCAGAACGAGCCATGTCCCTGTACGATTACTACTATTTCACAGAGAG	3038
Db	201876	CCTACTCTCGGCCCAGAACGAGCCATGTCCCTGTACGATTACTACTATTTCACAGAGAG	201817
QY	3039	AAGAGCCCAGGAGGTGGAAGGAGCGCTTTTGGCCGACAGGCTGCATCCCCCATCCAAATC	3098
Db	201816	AAGAGCCCAGGAGGTGGAAGGAGCGCTTTTGGCCGACAGGCTGCATCCCCCATCCAAATC	201757
QY	3099	ATGACGGTGTCAACATCTGCAGCTCCCACCTGAAATCGTGTCTCTCTGAAATCTCAGGAA	3158
Db	201756	ATGACGGTGTCAACATCTGCAGCTCCCACCTGAAATCGTGTCTCTCTGAAATCTCAGGAA	201697
QY	3159	GTGCCTATGGGAAGGACTATCCTCAAAGTCAACCCCGGAAAAACAACCTGTTCCAGCCCCC	3218
Db	201696	GTGCCTATGGGAAGGACTATCCTCAAAGTCAACCCCGGAAAAACAACCTGTTCCAGCCCCC	201637
QY	3219	GTGCGGAAGTACAACTCCAATGTCTAATATCATCACCGAAGACAATAAATTCACATT	3278
Db	201636	GTGCGGAAGTACAACTCCAATGTCTAATATCATCACCGAAGACAATAAATTCACATT	201577
QY	3279	CACCTGGTTCAGTTTAAAGCGATCTCTCGGGCTGCCGCTGAAGGCGTGAGCCCCAGTT	3338
Db	201576	CACCTGGTTCAGTTTAAAGCGATCTCTCGGGCTGCCGCTGAAGGCGTGAGCCCCAGTT	201517
QY	3339	ATCACCGTCCGGCTGTCAACGTGACAGCGGAGAAAGGAGTTTCTACAGGCACAGTCCCT	3398
Db	201516	ATCACCGTCCGGCTGTCAACGTGACAGCGGAGAAAGGAGTTTCTACAGGCACAGTCCCT	201457
QY	3399	CGCTCTCCAGGAACCACTCTCTTCAAGACCCCGGTGCTAGCAAAGTGACCAGCACTATA	3458

Db 201456 CGCTCTCCAGGAACCACTCTCTTCAAGACCCGGTGTAGCAAAAGTGACCAAGCACTATA 201399

QY 3459 ACTATAACCCCGGTACAAAGTCCATCCACACGAGGAACCCCAATCAGTG 3506

Db 201396 ACTATAACCCCGGTACAAAGTCCATCCACACGAGGAACCCCAATCAGTG 201349

RESULT 8
 LOCUS AB033101
 DEFINITION Homo sapiens mRNA for KIAA1275 protein, partial cds.
 ACCESSION AB033101
 VERSION AB033101.1 GI:6331293
 KEYWORDS .
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (sites)
 Nagase,T., Ishikawa,K., Kikuno,R., Hirosawa,M., Nomura,N. and
 Ohara,O.
 TITLE Prediction of the coding sequences of unidentified human genes. XV.
 The complete sequences of 100 new cDNA clones from brain which code
 for large proteins in vitro
 JOURNAL DNA Res. 6 (5), 337-345 (1999)
 MEDLINE 20039619
 PUBMED 10574462
 REFERENCE 2 (bases 1 to 4051)
 AUTHORS Ohara,O., Nagase,T. and Kikuno,R.
 TITLE Direct Submission
 JOURNAL Submitted (04-OCT-1999) Osamu Ohara, Kazusa DNA Research Institute,
 Laboratory of DNA Technology; 1532-3 Yana, Kisarazu, Chiba
 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,
 URL:http://www.kazusa.or.jp/huge/, Tel:+81-438-52-3913,
 Fax:+81-438-52-3914)

FEATURES
 source location/Qualifiers
 1. .4051
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="hk10413"
 /tissue type="brain"
 /clone_lib="pBluescriptII SK plus"
 1. .4051
 /gene="KIAA1275"
 <30. .3452
 /gene="KIAA1275"
 /note="Start codon is not identified."
 /codon_start=1
 /product="KIAA1275 protein"
 /protein_id="BAA86589.1"
 /db_xref="GI:6331294"
 /translation="AQAIVIVSRLLHLFLQAREDVIHMLKTEKTPKEVLEAHYGS
 AEPEKLVRLHRDAIIAQEKSIGEDVYEKPISELDRIEEKQKETYRMLEQLLLAEKC
 HRTVVELENEKHKHTDYNKSDDDFTNLLEQERERLKLLEQEKAYQARKEKENAKRL
 NKRLDELVLKSFALMLVDQRMHIEQLGLSQKVDLTQKLREEEKLKATSKSE
 DRQLKLVDLFEFKASRFQSEHEEMNAKLANOESHNRQLRLKLVGTQRIEELRETN
 KNLQAEELQLRDKIAKGECSNLSMAEVENLRKRVLEMEGKDEETKTESQCREL
 RKLQGEHHSEKLRLEVEKLRKMSLEKLEAFSKSECTQLHLNLEKEKNLTKD
 LLNELEVYKSRVKELECSERLKAELSLLKDDLTKLKSFVTLVDERKNMMEKIKQEE
 RKVDGLNKNFKVEQKQVMDVTEKLIERSKKLLKLKSGMEEKVYNLTRERDELIGKES
 EEKSELCSVDLLKKRLDGIIEVEEITRGRSRKGSSELTCPEDNKIKELTLEIRL
 KKRLQQLVEVGEDLMKTEDEYDQLEQKFTQDKANFLSQOEEIKHQIAKNKAIKGG
 EVVSQLAEHLRFRLEAKSRDLKAEVQALKKEKIHMLNKKEDQLSQLQVDYSVLQORF
 MEEENKNKMGQEVNLTKELELSKRYSRALRPSVNGRRMVDVPVTSVGQTDVSGE
 AAEEETPAVPIRKSFQENHIMSNLRQVLKKPVRSSVLDRYPAAENLTMRKSWIP
 WMKRENGPFIQEKGPRTNSSPGHGEVWLSPKQGQPLHIRTVPDHNSTATLEITS
 PTSEEFSSSTTVITLGNQKPRITIIIPSNVMPQKQSGDTTLGERAMSPVITITFS
 REKTPESGRGAFADPTSPQIMHTVSTSAAPAEIAVSPESQEMPGRITILKVTPEKQT
 VPTPVKRYSNANIIITEDNKIHIHLGSOFKRSPGTSGEVSPVITVRPVNVTAEKEV
 STGTVLRSPRNHLSPGASKVTSTIITPTVTSARGTQSVSGQDSSQRPPTPTPIP
 MSXGMKAGKPVVAAPGAGNLTKEFRAETQSMKIELKSAASSTISLGGKG"

gene
 CDS

Db 3226 CTCGAGGAACCCAGTCAGTGTCTCAGGACAAGACGGGTCTATCCAGCGGCTACACCCACCC 3285
QY 3547 GCATTCCTATGTCAAAAGGTAATGAAGCTGGAAGCCAGTAGTGGCAGCCTCAGGAGCAG 3606
Db 3286 GCATTCCTATGTCAAAAGGTAATGAAGCAGGAAGCCAGTAGTGGCAGCCCCAGGAGCAG 3345
QY 3607 GAAATCTGACCAAAATTCAGCCTCGAGCTGAGACTCAGTCTATGAAAAATAGAGCTGAAGA 3666
Db 3346 GAAATCTGACCAAAATTCGAGCCTCGAGCTGAGACTCAGTCTATGAAAAATAGAGCTGAAGA 3405
QY 3667 AATCTGCAGCCAGCAGCACTGCCTCTCTTGGAGGGGGGAAGGCTGAGGGCAGTGGCTAA 3726
Db 3406 AATCTGCAGCCAGCAGCACCACCTCTCTCGAGGGGGGAAGGCTGAGGGCAGTGGCTAA 3465
QY 3727 GGGGGTATGTTGTAAGGATGCTACTGCTGCGAGTGAAGAAACAAACCTTCTCTGTGCCAACC 3786
Db 3466 GGGGGTATGTTGTGCAGATGCTACTGCTGCCGTGAAGAGTGAACCTTCATCTGTTGTGCC 3525
QY 3787 CTTTCTTTGTA-CTACTAATTTAAGTTTAAATATCTTGTATATCTTGTATATAAAATAACCATTTAAT 3845
Db 3526 AGTCTTTACATGTAATAATTTAAGTTTAAATATGTTTATATAAAATAACCAACTAAT 3585
QY 3846 AGCCATGCACCCCCCTCCCAATTTGTGCACTCTGTTTCAATGCGAGGGGAATAGAATTAAT 3905
Db 3586 AACCATT--TGTCTTTCCCAATTTGTGCAATTTGTTGATGCTGGGAACAA---AATT 3639
QY 3906 AGCAGAAATTTCTGTTGCTGAATGTTCTGTTGAAGATGTTGGTCCAGTTCAGTTTAACTT 3965
Db 3640 AGCAAAACTATTGCTTGCTGCTAGAAGCCAGGGCGTGGTTCTAGTTCAGTTTGGCTT 3699
QY 3966 CTAGCATGTGCCCCCAATTCAGGCTAGCTCAGCAGTGTGTAAGCCCTCAATATCGTCACCG 4025
Db 3700 CTAGCAAGTGGACCCATCAATAGACCCATCTGAG-----CCTGTTTCTCTCATCA 3748
QY 4026 GAGAGATTTGAGGACCACATTAACATATGCTCCCAAAGGCTGGCTCCCAATTTTCCTAAT 4085
Db 3749 GTTAGATGTGGGACTCAAT--CACACGCTCTTCAAGTCGGCTCCCAATTTTCCTAAT 3806
QY 4086 GTAAGCCCACTTTAATAGACTCAGTTCTGTGAT-TTTTTTTTCCAAAAAAATATTTT 4144
Db 3807 GCAAGCCCAATTTAATGTACTTGTTCACAATAATTTTATTAATAAAATCCTATTAC 3866
QY 4145 GAAATAGACAGAGTTTAAAGTTGTCAATTTGCACTATCAAGCCATGAGTTTGATATAT 4204
Db 3867 AAAATAAGACATACTTTAACTATTGTCAATTTGCCTCTTTCACATCATGAATTTGCTTTAT 3926
QY 4205 GGGTTATAAGAAA-AGAATACTTTTCAGAGCTATCACAGGCTCTCTAAACTTTTGGRAAAA 4263
Db 3927 GTGCTGAAAAAACAATCACAATAGCTATCACAGGCGCTTGGAACCTTAATAATTTTGCAAAAA 3986
QY 4264 CAAAAGCCCTAATATGACCTCAGGAACAATTTGAACATGAAATAAAATGGAATGAAC 4323
Db 3987 CAAAAGGTTCTAAGATGATTTTCAGGAATAAATGTGAACATGTAATAAAATGGAATGAAA 4046
QY 4324 TGTGG 4328
Db 4047 TATGG 4051

RESULT 9
AX780844
LOCUS AX780844 4065 bp DNA linear PAT 14-JUL-2003
DEFINITION Sequence 3001 from Patent WO03039443.
ACCESSION AX780844
VERSION AX780844.1 GI:32697838
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Haerlach,T., Schoch,C., Kern,W., Kohlmann,A., Schnitger,S.,
Dugas,M., Eils,R., Brors,B. and Mergenthaler,S.

TITLE Novel genetic markers for leukemias
JOURNAL Patent: WO 03039443-A 3001 15-MAY-2003;
Deutsches Krebsforschungszentrum (DE);
Ludwig-Maximilian-Universitaet Muenchen (DE); Haerlach, Torsten,
PD Dr. Dr. (DE); Schoch, Claudia (DE); Kern, Wolfgang (DE)
FEATURES
source Location/Qualifiers
1..4065
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 63.2%; Score 2760.2; DB 6; Length 4065;
Best Local Similarity 81.9%; Pred. No. 0;
Matches 3309; Conservative 0; Mismatches 704; Indels 26; Gaps 11;
QY 307 ACCTCATCCAGCTCCTGAGTATCATGGAAGGGGAGTTGCAGGGCTCGAGAAGATGTCATCC 366
Db 46 ACATTGTCAGTCGTCATCTTGTGCACTTTCTATTACAGGCCAGAGAAGATGTGATCC 105
QY 367 ACATGCTGAGGACAGAGAAAACCAAGCCCGAGGTTCTGAGGGCACACTATGGATCTGCAG 426
Db 106 ACATGCTGAAGACAGAGAAAACCAAGCCTGAGGTTCTGAGGCTCATTACGGGTCTGCG 165
QY 427 AACCTGAGAAAGTGCTTCGGGTCTTGACCCGAGATGCCATCCTTGTCAAAGAAAGTCCA 486
Db 166 AGCCAGAGAAAGTGCTGCGGTCTTGACCCGAGATGCCATTTTGCCAGGAGAAATCCA 225
QY 487 TAGGAAAGACGCTCTATGAGAAAACCTTATCTCAGAGCTGGACAGACTGGAGGAAAAGCAGA 546
Db 226 TAGGAGAAGATGTCATGAGAAACCGATTTTCAGAGCTGGACAGACTTGAGGAAAAACAGA 285
QY 547 AGGACACGTACCGCCGCATGCTAGAGCAGCTGCTGCTGGCTGAGAGTGTTCACAGGGCCA 606
Db 286 AAGAAACCTACCGCGCATGCTAGAGCAGCTGTTGCTGGCCGAGAGTGTTCATAGCGCA 345
QY 607 CCGTGTACGAGCTGGAGAACGAGAAACACAGACACACTGACTACATGAACAAGAGCGACG 666
Db 346 CCGTNTCGAGTTAGAGAACGAGAAAGCATATAACACACTGACTACATGAACAAGAGCGACG 405
QY 667 ACTTCACCAACCTGCTGGAGCAGGAGCGAGAGGTTGAAAAAGCTCCTTGAACAAGAAA 726
Db 406 ACITCACCAACCTGCTGGAGCAGGAGCGGAGAGGTTAAAAAGCTCCTTGAACAAGAAA 465
QY 727 AAGCTTACCAAGCCCGCAAGAAAAGGAAAAACCGCTAAAGCGGCTCAACAACTTCGAGATG 786
Db 466 AGCTTATCAAGCCCGCAAGAAAAGGAAAAATGCTAAAACGACTCAATAAATAAGAGATG 525
QY 787 AGCTTGTGAAGCTCAAGTCCITTCGCCCTCATGTTGGTGGACGAGAGGCAGATGCACATCG 846
Db 526 AGCTTGTAAACTCAAATCCTTTGCACCTCATGCTGGTGGATGAAAAGACAAATGCACATTG 585
QY 847 AGCAACTGGGCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAAGCTGAGGAGGAGG 906
Db 586 AACAACTTGGCCTGCAAGCCAGAAAGTACAGGATCTTACTCAGAAAGCTGAGGGAAGAAG 645
QY 907 AAGAAAAACTCAAAGCGGTCACTTACAAATCCAAAGGAAGACCGCCAGAAAGCTGCTCAAGT 966
Db 646 AAGAGAAGCTCAAAGCCATTACTTCCAAATCCAAAGAGACAGACAGAAATTGCTCAAGT 705
QY 967 TAGAAGTGGACTTCGAACACAAGGCCTCGAGGTTTTTCCCAGGAGCACGAGAGATGAACG 1026
Db 706 TAGAAGTGGACTTTGAACACAAGGCTTCGAGGTTTTCTCAAGAGCATGAAGAGATGAACG 765
QY 1027 CCAAATTGGCGAATCAAGAAATCTCACAAACCGGCAACTTCGACTCAAACTGGTTGGCTTAT 1086
Db 766 CTAAACTGGCTAATCAAGAGTCTCAATATAGGCAACTTAGACTCAAGCTGGTTGGCTTAA 825
QY 1087 CGCAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAAGCAGAGGAAGAGC 1146
Db 826 CCCAAAGAAATCGAGGAGCTAGAAGAGACCAACAAAATCTGCAGAAAGGCAGAGGAAGAAC 885
QY 1147 TCCAGGAGCTGAGAGAGAAAAAATTGCCAAAGGGGAATGTGGAAACTCCAGTCTCATGGCGG 1206

Db	886	TTCAAGAATTAAAGAGATAAAATTCGCAAGGAGAATGTGGAACCTCTAGCCCTCATGGCAG	945
QY	1207	AACTGGAGAGTCTGCGCAAGCGGTGCTTGAGATGGAGGGCAAGGATGAAGAGATCACGA	1266
Db	946	AAGTGGAATACTTCGAAGCGGTGCTTGAAATGGAAGGTAAAGATGAGGAGTCACTA	1005
QY	1267	AGACCGAGGCCAGTCCCGGAGCTGAAGAAGAGCTCCAAGAGGAAGAACACCCACAGCA	1326
Db	1006	AAACTGAATCCCAGTGTAGGGAATTGAGGAAGAGCTGCAAGAGGAAGAACACCATAGTA	1065
QY	1327	AGGAACCTTAGACTAGAACTGGAGAAGCTGCAGAAAGAGATGTCTGAGCTGGAGAAGCTGG	1386
Db	1066	AGGAGCTCAGACTTGAAAGTTGAGAAGCTACAGAAGAGAATGTCTGAACCTAGAGAAATTGG	1125
QY	1387	AGGAAGCGTTTCAGCCGGAGTAAAGTCGGAATGCACCCAGTCCATCTGAACCTGGAGAAGG	1446
Db	1126	AAGAAGCATTTAGCAAGAGTAAATCTGAGTGCACCCAGCTACATTTTAAATCTGGAGAAAG	1185
QY	1447	AGAAGAACCTTAACCAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAG	1506
Db	1186	AAAAGAACTTAACCAAGACCTGCTAAATGAATTTGGAGGTGGTCAAGAGTCGAGTTAAAG	1245
QY	1507	AACCTCGAATGCTCCGAGAGTAGACTGGAGAAGGCCGAGTTAAGCCTCAAAGATGACCTTA	1566
Db	1246	AATTGGAATGTTCTGAAAGTAGATTGGAAAAGGCTGAATTAAGCCTAAAAGATGATCTTA	1305
QY	1567	CAAAGCTGAAGTCCTTCACTGTGATGCTGGTGGATGAGAGGAAAAATATGATGAGAAAA	1626
Db	1306	CCAAGTTGAAGTCATTTACCGTGATGCTGGTTGATGAAAGGAAAAATATGATGGAAAAA	1365
QY	1627	TAAAGCAAGAAGAGAGGAAAGTGGATGGTTGAATAAAAACTTTAAGGTGGAGCGGAA	1686
Db	1366	TAAAAACAAGAGAGAAAAAGTGGATGGACTCAATAAAAAATTTTAAGGTGGAAACAAGGA	1425
QY	1687	AAGTCATGGATGTGACGGAAAAAGCTAATCGAGGAAAGCAAGAAGCTTTTAAAACTCAAAT	1746
Db	1426	AAGTTATGGATGTAACCTGAAAAAATAATTGAAGAAAGTAAGAAACTTTTAAAACTAAAT	1485
QY	1747	CTGAATGGAGGAAAAAGGAGTACAGTCTGCACAAAAGGAGGGATGAGCTGATGGTTAAAC	1806
Db	1486	CTGAATGGAGGAAAAAGTATACAACCTTGACAAGAGAAAGAGATGAGTTGATAGGCAAT	1545
QY	1807	TGAGGAGCGAAGAAGGTCCTGTGAACCTGAGCTGCAGTGTAGACTTACTTAAAGAAGC	1866
Db	1546	TGAAGAAGTGAAGAAGAAAAATCCTCTGAATTAAGCTGCAGTGTGACTTTACTTAAAGAAG	1605
QY	1867	GGCTTGATGGCATAGAGGAGGTAGAAAGGGAAATAAACCGAGGTAGGTGTCGAAGGGGT	1926
Db	1606	GACTTGATGGTATAGAGGAAGTGGAAAGAGAAATAACAAGAGGAAGGTACGAAAAAGGGT	1665
QY	1927	CTGAGTTCACTGCCCGAAGACAAATAAGATCAGAGAACTAACGCTTGAAATCGAGAGAC	1986
Db	1666	CTGAGCTCACCTGCCCGGAAGATAATAAGATTAAAGGAATAACACTTAACACTTGAAATTGAGAGAC	1725
QY	1987	TGAAGAAACGGCTCCAGCAGTTGGAGGTGGTGGAGGGGACTTGATGAAGACCGAGGACG	2046
Db	1726	TGAAGAAACGTCCTCCAACAATTGGAAGTGGTTCGAAGGGGATTTGATGAAGACAGAAGATG	1785
QY	2047	AATATGACCAGTTGGAGCAGAAAGTTCAGAACCGGAGGAGGATAAGGCAAACTTCTCTCCC	2106
Db	1786	AGTATGATCAGCTGGAACACAGAAATTTAGAACTGAGCAGGATAAAGCTTAACCTTCTCTCTC	1845
QY	2107	AGCAGCTCGAGGAAATCAAACACCAAATGGCCAAAGCACAAGGCCATAGAGAAAGGGGAGG	2166
Db	1846	AACAACTAGAGGAGATCAAGCACCAAATTGCCAAGATAAAGCAATAGAGAGGGTGAGG	1905
QY	2167	CCGTGAGCCAGGAAGCCGAACCTGCGACACAGGTTTCGGCTGGAGGAGGCTAAAAGTCGTG	2226
Db	1906	TTGTGAGCCAGGAAGCTGAACCTGAGACACAGATTTTCGGTTGGAAGAAGCTAAAAGTCGAG	1965
QY	2227	ATTTACAGGCCGAGGTGACGGCTCTCAAGGAGAAGATCCACGAGCTGATGAACAAGGAAG	2286

Db	1966	ACTTAAAGCCGAAGTACAAGCTCTTTAAAGAGAAAGATTACGAATTAATGAACAAAGAAG	2025
QY	2287	ACCAGCTGTCTCAGCTCCAAGTCGACTATTCCGTCTTTCAGCAAAGATTTTATGGAAGAAG	2346
Db	2026	ATCAGCTTTCTCAGCTCCAGGTAGATTATTCTGTACTTCAACAAAGATTTTATGGAGAAG	2085
QY	2347	AAACTAAGAACAAAGACATGGGGAGGGAGGTCTCTCAATCTTGACCAAGGAGCTAGAGCTTT	2406
Db	2086	AAAATAAGAACAAAAACATGGGCGAGGAGTTCTCAATCTGACCAAGAGTTGGAGCTTT	2145
QY	2407	CAAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGTGGAAACGGCCGAAGGATGGTGGACGTGC	2466
Db	2146	CAAAGCGCTACAGCAGAGCTCTTAGGCCCAGTGTGAATGGAAGAAGATGGTGGATGTTTC	2205
QY	2467	CTGTGGCCTCCACTGGGGTGCAGACCCGAGGCGGTGTGCGGGATGCTCGGGAGGAGAGA	2526
Db	2206	CTGTGACGTCAACTGGAGTCCAAACTGATGCACTCAGCGGTGAAGCAGCAGAGGAAGAA	2265
QY	2527	CCCCGGCTGTGTTCAATTCGCAAAATCCTTCCAGGAGGAAAAATCACATCATGAGTAATCTTC	2586
Db	2266	CGCCAGCTGTATTCAACGGAATCCTTCCAGGAAGAAAAATCATATTATGAGTAATCTTTC	2325
QY	2587	GACAGGTAGGCTGAAGAAAAACCATGGAACGGTCTCGGTCTCTCGACAG-GTATCCCCCA	2645
Db	2326	GGCAGGTGGATTGAAGAAAAACCGTGGAAAGATCTTCTGTCTAGACAGNGTATCTCTCA	2385
QY	2646	GCAGCGAATGAGTCAACCATGAGGAAGTCTTGGATTCTTGGATTCCCTGGATGAGAAAAAGAAAAAC	2705
Db	2386	GCAGCAAAATGAGTCACTATGAGAAAGTCTTGGATTCCATGGATGAGAAAGAGGGAAGAAAC	2445
QY	2706	GGTCTTCCACTCCGAGGAGAAAGGGCCCCAGSCCAACACAGGGTGCAGGGCACCCTGGG	2765
Db	2446	GGCCCNCTCCATCACTCAGGAGAAAGGGCCCCCGAACAAATTCAGTCCAGGGCACCAGGA	2505
QY	2766	GAGCTGGTCTTAGCACCAAAGCAGGGCCAGCCCTACACATCCGTGTGACACAGATCAT	2825
Db	2506	GAGGTAGTCTTTTCAACCAAGCAGGGCCAGCCCTGCATATTCGAGTGACACACAGACCAC	2565
QY	2826	GAGAACAGCAGCTGCCACCTGGAGATCACAGCCCCACATCTGAAGAGTTTTTCTCTAGT	2885
Db	2566	GAGAACAGCAGCTGCGACTTTGGAGATAACAAGCCGACATCTGAAGAAATTTTTTCTAGT	2625
QY	2886	ACCACCGTCATTCTTACCTTAGGCAACCAAGAACCAAGAAATAACCATATTATCCATCACCC	2945
Db	2626	ACCAGTGTCAATCTCTACCTTAGGGAATCAGAAACCAAGAAATAACCATATTATCCATCACCA	2685
QY	2946	AATGTCATGTGCGAAAAAGCCCCAAAAGTGCAGATCTCTCTCGGCCCAGAACGAGCCATG	3005
Db	2686	AACGTTATGCCTCAAAAACAAAAGTGGAGATACTACTCTTGGCCCCAGAACGAGCCATG	2745
QY	3006	TCCCTGTCAAGATTACTACTATTTTCCAGAGAG-AAGAGCCCCGGAAGGTGGAAGGAGCGC	3064
Db	2746	TCCCCAGTCACAATTACTACATTTTCCAGAGAGNAAGACTCCAGAAAAGTGGAAAGGCGC	2805
QY	3065	CTTTGCCGACAGGCTGCATCCCCCATCCAATCATGACGGTGTCAACATCTGCAGCTCC	3124
Db	2806	ATTTGCAGACAGGCCACATCCCCTATTTCAGATAATGACGGTGTCTACATCAGCAGCAC	2865
QY	3125	CAGTGAATCGTGTCTCTCTGAATCTCAGGAAGTGCCTATGGGA-GGACTATCCTCA	3183
Db	2866	AGCTGAGATTGCACTTCTCCCGAATCCCAGGAAATGCCCATGGNACGGACAATCCTCA	2925
QY	3184	AAGTCACCCCGGAAAAACAAACTGTTCCAGCCCCCGTCCGGAAGTACAACTCCAATGCTA	3243
Db	2926	AAGTCACCCCGGAAAAACAGACNGTTCCAACTCCAGTACGGAAATACAACCTCCAATGCCA	2985
QY	3244	ATATCATCACCCGGAAGACAATAAAATTCACATTCACCTGGGTCTCAGTTTAAAGCGAT	3303
Db	2986	ATATCATAAACACAGAGGACAATAAAATTCACATTCACCTTAGGGTCTCAGTTTAAACGGT	3045
QY	3304	CTCCTGGGCTGCCGCTGAAGGCGTGAGCCAGTTATCACCGTCCGGCTGTCAACGCTGA	3363
Db	3046	CCCCTGGACTTCAGGTGAAGGAGTCACTCCAGTTATTACTGTCCGACCAGTAAACGCTGA	3105

QY	3364	CAGCGGAGAAAGGAGGTTTCTACAGGCACAGTCCTTCGCTCTCCAGGAACCACTCTCTT	3423
Db	3106	CAGCCGAAAAGGAGGTTCCACGGCACTGCTTCCTCTCCAGGAATCACCTCTCCT	3165
QY	3424	CAAGACCCGGTGTAGCAAAAGTGACCAAGCTGACGACACTATAACTATAACCCCGGTACAAAGTCAT	3483
Db	3166	CACGGCTGGTGCAAGCAAAAGTGACGAGCACTATCACCATAACACCGGTACAAAGTCAT	3225
QY	3484	CCACAGGAGAACCCCAATCAGTGTGACGACAAAGATGGGTCTATCTCAGCGGCCTACCCCCA	3543
Db	3226	CTGCTCGAGGAACCCAGTCAGTGTGAGGACAAAGACGGGTCTATCCAGCGGCCTACACCCA	3285
QY	3544	CCCGCATTCTATGTCAAAAGGTATGAAAGCTGGAAGCCAGTAGTGCGAGCCTCAGGAG	3603
Db	3286	CCCGCATTCTATGTCAAAAGGTATGAAAGCAGGAAAGCCAGTAGTGCGAGCCCGCAGGAG	3345
QY	3604	CAGGAAATCTGACCAAAATTCAGCCTCGAGCTGAGACTCAGTCTATGAAAATAGAGCTGA	3663
Db	3346	CAGGAAATCTGACCAAAATTCAGCCTCGAGCTGAGACTCAGTCTATGAAAATAGAGCTGA	3405
QY	3664	AGAAATCTGCAGCCAGCAGCACTGCCTCTCTTGAGGGGGGAGGGCTGAGGGCAGTGGC	3723
Db	3406	AGAAATCTGCAGCCAGCAGCACCACCTCTCTCGAGGGGGGAGGGCTGAGGGCAGTGGC	3465
QY	3724	TAAGGGGGTATGTTGTAAGGATGCTACTGCTGAGTGGAACCAACCTTCCTCTGTGCCA	3783
Db	3466	TAAGGGGGTATGTTGTCAGATGCTACTGCTGCGTGAAAGTGAACCTTCACTGTNNGT	3525
QY	3784	ACCCTTTCTTGTA-CTACTAATTTAAGTTTTPAAATATCTTGTTTATA-AAATAACCAAT	3841
Db	3526	GCCAGTTCTTACATGTACTAATTTAAGTTTTPAAATATNGTNTATANATATAACCAAC	3585
QY	3842	TAATAGCCATGCACCCCCCTCCCATTTTGTGATCTGTTTCAATGCAGGGGAATAGAAT	3901
Db	3586	TAATAACCAT--TNGTCTTTCCCATTTNGTGCATTTGTTINGATGCTGGGAACA---N	3639
QY	3902	AATTAGCAGAAATTTCTGTTTGCTGAATGTTCTGTTGAAGATGTTGGTCCAGTTCAGTTTT	3961
Db	3640	AATTAGCAAAACTATTGCTTGCTGCCTAGAACCCAGGCGTGGTTTCTAGTTCAGTTTT	3699
QY	3962	ACTTCTAGCATGTGGCCCCCATTCAGGTAGTCCAGAGTGTGGAAGCCCTCAATATCGTC	4021
Db	3700	GCTTCTAGCAAGTGGACCCCATCAATAGACCATCTGAGC-----CTGTTTCCCTC	3748
QY	4022	ACCGGAGAGATTTGAGGACCACATTAACATATGCTCCCAAGGCTGGCTCCCAATTTTCCT	4081
Db	3749	ATCAGTTAGATGTGGGACTCAATCACAC--GCTCTTCAAGTCCGGCTCCCATATTTCTC	3806
QY	4082	AATTGTAAGCCAACTTTAATAGACTCAGTTCTGTGAT-TTTTTTTTCCAAAAAATAA	4140
Db	3807	AATTGCAAGCCAAATTTAATGTACCTTGTTCCACAATAATTTTTTATTAATAAATCCTA	3866
QY	4141	TTTTGAAATAGGACAGAGTTTAAACAGTTTGCATTTTGCACTATCAAGGCTCTCTAAAC	4200
Db	3867	TTACAAAATAAGACATACCTTTAACTATTGCTCATTTGCTCTTTTACATCATGAATTTGCT	3926
QY	4201	ATATGGGTTATAAGAAA-AGAAATCTTTTCAGAGCTATCACAGGGTCTCTAAAC	4259
Db	3927	TTATGTGCTGGAAAAACATCATAGCTATCACAGGCTTGACCTCTAAAAATTTTGCA	3986
QY	4260	AAAACAAAAGCCCCCTAATATGACCTCAGGAAACAAATTTGAACATGAATAAATGGAAAT	4319
Db	3987	AAAACAAAAGGTTCTAAGATGATTTTCAGGAAATAATGTGAACATGTATAAATGGAAT	4046
QY	4320	GAACGTGGAATCTTAAAA 4338	
Db	4047	GAATAATGGAATCCTAAAA 4065	

RESULT 10
HSM803316
LOCUS HSM803316 4793 bp mRNA linear PRI 17-JUN-2003

DEFINITION Homo sapiens mRNA; cDNA DKFZp451B134 (from clone DKFZp451B134); complete cds.
ACCESSION AL832009
VERSION AL832009.2 GI:30268230
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4793)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2003) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
COMMENT On Apr 30, 2003 this sequence version replaced gi:21732549. Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFZp451B134) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cDNA/.
FEATURES
Location/Qualifiers
1..4793
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp451B134"
/db_xref="taxon:9606"
/map="6q14.1"
/clone="DKFZp451B134"
/tissue_type="human skeletal muscle"
/clone_lib="451 (synonym: hlcc1). Vector pSport1; host DH10B; sites NotI + SalI"
/dev_stage="adult"
1..4793
/gene="DKFZp451B134"
223..3756
/gene="DKFZp451B134"
/note="similar to filamin-interacting protein L-FILIP (Rattus norvegicus)"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD89912.1"
/db_xref="GI:30268231"
/translation="MRSRNGGESASDGHISCPKPSIIGNAGEKSLSEDAKKRKSNR KEDDVMAAGTGVKRLTKSGECERKTKSLELSKEDLIQLLSIMEGELOAREDVHMLK TEKTKEVLEAHVGSAPPEKVLVLRDAILAQEKSIGEDVYEKPISELDREEKQKE TYRMLQQLLAEKCHRTTYELENEKHGHTDYMNSDDFTNLLQERERLKLLEQE KAYQARKEKENAKRLNKLDELVLKLSFALMLVDEROMHIEQLGLSQKVQDILTQKL EEEKLKAITSKSKEDQKLKLEVEFHKASRFSQHEEMNAKLANOESHNRQLRK LVGLTORIEELEENKLNQAAEEELQELRDKIAKGECNSSLMAEVENLRKRVLEMEG KDEITKTESQCRELRKKLQEEHHSKELRLEVEKLQKRMSELEKLEAFSKSSECT QLHLNLEKKNLTKDLENEVSVKVELECSERLKAELSLKDDLTCLKSFTVML VDERKNMMEKI KOERKVDGLKNFKVQGKMDVTEKLIIEESKLLKLKSEMEKCV NLTRERDELIGLKSEESSELSVLDLKKRLDGTBEVEREITRGRSKGSELTCP EDNKTKELTLEIERLKRLLQQLLEVVEGDLMTKTEYDQLEQKFRTEODKANFLSQOLE EIKHOIAKNKAIIEKGVVVSQAEALHRFRLEEAKSRLKAEVQALKKEIHELMNKEDQ LSQLDVSVLQQRFEENKKNMGQEVNLNLTKELELSKRYSRALRPSVNGRRMVDV PVTSTGVTDAVSGEAAEEETPAVFIRKSFOEENHIMSNLRQVGLKRPVERSSVLDRY PPAANELTMRKSWIPWMRKRENGPSITQKGPRTNSPFGPGEVLSPKQGQPLHIRV TPDHENSTATLEITTFSEKTPESGRGAFADPTSPIQIMTVSTSAAPAEIAVSPESQEM GPERAMSPVTITTFSEKTPESGRGAFADPTSPIQIMTVSTSAAPAEIAVSPESQEM PMGRTILKVTPEKQVPTPVKKNYNANIITTEDNKIHLGSLQKRSFGTSGEVSP VITVPVNVNTAEKGVSTGTVLSPRNHLSSRPGASKVTSTITITPVTTSARGTQSVS QDGSQRPTPTTRIPMSKESIIHQLRMNSR"
4773
/gene="DKFZp451B134"
polyA_site
ORIGIN
Query Match 60.5%; Score 2641; DB 9; Length 4793;

Best Local Similarity 84.4%; Pred. No. 0; Matches 2982; Conservative 0; Mismatches 549; Indels 4; Gaps 2;									
QY	34	AAAGAACACACACATTGGTT-AAGGAGTCGACAAACAGGTGGGAATGAGATCACGAAATCAA	92						
Db	181	AGAGGACATACCTTTGGTTAAAGGAGCTGCCGTGAGTGGGAATGAGATCTCGAAACCAA	240						
QY	93	GGTGGAGAAAGTTCATCTAACGGGCATGTCTCCTGCCCCCAAGTCCTCCATCATCAGCAGT	152						
Db	241	GGTGGTAAAGTGATCTGATGGGCATATCTCCTGTCTCCCAAGCCCTCCATCATCGGCAAT	300						
QY	153	GATGGTGGTAAGGGCCCTCAGAAGATGCAAAAAAGAA---CAAGGCCAATCGGAAGGAG	209						
Db	301	GCTGGTGAAAAAAGTCTCTCAGAAGATGCAAAAAAGAAAGAGGAAATCAAATAGGAAGGAG	360						
QY	210	GAGGATGTATGGCTTCCGGAATATCAAAAGGCACCTCAAAACCATCTGGAGAAAGTGAG	269						
Db	361	GATGATGTATGGCCTCAGGAAGTGTCAAAACGACACCTAAAAACATCTGGAGAATCTGAA	420						
QY	270	AAAAAGACTAAGAAAGTCTGTGGAGTTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATC	329						
Db	421	CGAAAAAACTAAGAAATCCCTGGAGTTATCCAAGAAAGACCTCATCCACTACTCACTATA	480						
QY	330	ATGGAAGGGGAGTTGCAGGCTCGAGAAGATGTCTATCCACATGCTGAGGACAGAGAAAACC	389						
Db	481	ATGGAAGGGGAGTTGCAGGCCAGAGAAGATGTGATCCAATGCTGAAGACAGAGAAAAACC	540						
QY	390	AAGCCCGAGGTTCTGGAGGCACACTATGGATCTGCAGAAACCTGAGAAAGTGCTTCGGGTC	449						
Db	541	AAGCCTGAGGTTCTGGAGGTCATTACGGGTCTGCGGAGCCAGAGAAAGTGCTGCGGGTC	600						
QY	450	CTGCACCGAGATGCCATCCTTGCTCAAGAGAAAGTCCATAGGAGAAAGACGTCTATGAGAAA	509						
Db	601	CTGCACCGAGATGCCATTCTTGCCAGGAGAAATCCATAGGAGAAAGATGTCTATGAGAAA	660						
QY	510	CCTATCTCAGAGCTGGACAGACTGGAGGAAAAAGCAGAAAGGAGACGTACCGCCGATGCTA	569						
Db	661	CCGATTTCAGAGCTGGACAGACTTGAGGAAAAACAGAAAGAAACCTACCGCGCATGCTA	720						
QY	570	GAGCAGCTGCTGGCTGAGAAGTGTACAGGCGCACCGTGCTACGAGCTGGAGAACGAG	629						
Db	721	GAGCAGCTGTTGCTGGCCGAGAAGTGTATAGGCGCACCGTATACGAGTTAGAGAACGAG	780						
QY	630	AAGCACAAAGCACACTGACTACATGAACAAGAGCGACGACTTCACCAACCTGCTGGAGCAG	689						
Db	781	AAGCATAAACACACTGACTACATGAACAAGAGCGACGACTTCACCAACCTGCTGGAGCAG	840						
QY	690	GAGCGAGAGAGGTTGAAAAAGCTCCTTGAACAAGAAAAAGCTTACCAAGCCCGCAAAGAA	749						
Db	841	GAGCGGGAGAGGTTAAAAAAGCTCCTTGAACAAGAAAAAGGCTTATCAAGCCCGCAAAGAA	900						
QY	750	AAGGAAAAACGCTAAGCGGCTCAACAAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCTTC	809						
Db	901	AAGGAAAAATGCTAAACGACTCAATAAACTAAGAGATGAGCTTGTAAACTCAAATCCTTT	960						
QY	810	GCCCTCATGTTGGTGGACGAGAGGCAGATGCACATCGAGCAAACTGGGCCCTGCAGAGTCAG	869						
Db	961	GCACTCATGCTGGTGGATGAAAGACAAATGCACATTGAACAACTTGGCCTGCAAAAGCCAG	1020						
QY	870	AAAGTCCAGGACCTCACTCAGAAGCTGAGGGAGGAGGAAGAAAACTCAAAGCCGTCACT	929						
Db	1021	AAAGTACAGGATCTTACTCAGAAGCTGAGGGAAGAAAGAGAAAGCTCAAAGCCATTACT	1080						
QY	930	TACAAATCCAAGGAAGACCGCCAGAAAGCTGCTCAAGTTAGAAGTGGAATTCGAACACAAG	989						
Db	1081	TCCAAATCCAAGGAAGACAGACAGAAAATTGCTCAAGTTAGAAGTGGACTTTGAACACAAG	1140						
QY	990	GCCTCGAGGTTTTCCAGGAGCACGAAGAGATGAACGCCAAAATTGGCGAATCAAGAATCT	1049						
Db	1141	GCTTCGAGGTTTTCTCAAGAGCATGAAGAGATGAACGCTAAACTGGCTAATCAAGAGTCT	1200						
QY	1050	CACAAACGGCAACTTCGACTCAAACTGGTTGGCTTATCGAAAGGATTGAGGAGCTGGAA	1109						

Db	1201	CACAAATAGGCAACTTAGACTCAAGCTGGTTGGCTTTAACCCAAAGAAATCGAGGAGCTAGAA	1260
QY	1110	GAGACCAATAAAAAGCCTTCAGAAAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAATT	1169
Db	1261	GAGATCAACAAAAATCTGCAGAAAGGCAGAGGAAGAACTTCAAGAAATTAAGAGATAAAATT	1320
QY	1170	GCCAAAGGGGAATGTGAAAACTCCAGTCTCATGTCGCGGAAGTGGAGAGTCTGCGCAAGCGC	1229
Db	1321	GCCAAAGGAGAAATGTGAAAACTCTAGCCTCATGGCAGAAAGTGGAANAATCTTCGAAAGCGT	1380
QY	1230	GTGCTTGAGATGGAGCAAGGATGAAGAGATCACGAAGACCGAGGCCAGTCCCGGAG	1289
Db	1381	GTGCTTGAAATGGAAGTTAAAGATGAGGAGATCACTAAAACTGAATCCCAGTGTAGGAA	1440
QY	1290	CTGAAGAAAGAGCTCCAAGAGGAAGAACACCACAGCAAGGAACTTAGACTAGAAGTGGAG	1349
Db	1441	TTGAGGAAGAGCTGCAAGAGGAAGAACACCATAGTAAGGAGCTCAGACTTGAAGTTGAG	1500
QY	1350	AAGCTGCAGAAGAGGATGTCTGAGCTGGAGAGGCTGGAGGAAGCGTTTCAGCCCGAGTAAG	1409
Db	1501	AAGCTACAGAAGAGAAATGTCTGAACCTAGAGAAATTGGAAGAACATTTAGCAAGAGTAAA	1560
QY	1410	TCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAGGAGAAACCTTAACCAAGACCTG	1469
Db	1561	TCTGAGTGCACCCAGCTACATTTAAATCTGGAGAAAAGAAAGAACTTAACCAAGACCTG	1620
QY	1470	CTGAACGAGCTGGAGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGA	1529
Db	1621	CTAAATGAATTGGAGTGGTCAAGAGTCGAGTTAAAGAAATTTGGAATGTTCTGAAAAGTAGA	1680
QY	1530	CTGGAGAAAGCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAAGTCCCTTCACCTGTG	1589
Db	1681	TTGGAAAAGGCTGAATTAAGCCTAAAGATGATCTTACCAAGTTGAAGTCATTTACCGTG	1740
QY	1590	ATGCTGGTGGATGAGAGGAAAAATATGATGGAGAAAAATAAAGCAAGAGAGAGGAAAAAGTG	1649
Db	1741	ATGCTGGTTGATGAAGGAAAAATATGATGGAAAAATAAACAAGAGAGAGAAAAAGTG	1800
QY	1650	GATGGTTGAATAAAAAATTTTAAGGTGGAGCAGGAAAAAGTCAATGGATGTGACGGAAGAAG	1709
Db	1801	GATGGACTCAATAAAAAATTTTAAGGTGGAACAAAGGAAAAAGTTATGGATGTAACTGAAAAA	1860
QY	1710	CTAATCGAGGAAGCAAGAAGCTTTTAAAACTCAAAATCTGAAATGGAGGAAAAAGGAGTAC	1769
Db	1861	CTAATTGAAGAAAGTAAGAAAACTTTTAAAACTAAAAATCTGAAATGGAGGAAAAAGTATAC	1920
QY	1770	AGTCTGACAAAGGAGAGGGATGAGCTGATGGGTAAAACTGAGGAGCGAAGAAAAAGGTCC	1829
Db	1921	AAGTTGACAAAGAGAAAGAGATGAGTTGATAGGCAAAATTTGAAAAGTGAAGAAAGAAAAATCC	1980
QY	1830	TGTGAACCTGAGCTGCAGTGTAGACTTACTAAAGAAGCGGCTTGATGGCATAGAGGAGGTA	1889
Db	1981	TCTGAATTAAGCTGCAGTGTGACTTACTAAAGAAGAGACTTGTGATGGTATAGAGGAAGTG	2040
QY	1890	GAAAGGGAAATAAACGAGAGTAGGTGCTGTCAGGGGTCTGAGTTTACCTGCCCGGAAGAC	1949
Db	2041	GAAAGAGAAATAACAAGAGGAAGGTCAAGAAAAAGGGTCTGAGCTCACCTGCCCGGAAGAT	2100
QY	1950	AATAAGATCAGAGAACTAAACGCTTGAAATCGAGAGACTGAAGAAAAAGGTCCAGCAGTTG	2009
Db	2101	AATAAGATTAAAGGAATAACACTTGAAATTTGAGAGACTGAAGAAAAAGTCTCCAACAATTG	2160
QY	2010	GAGGTGGTGGAGGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAGCAGGAAG	2069
Db	2161	GAAGTGGTCGAAGGGGATTGATGAAGACAGAGAGATGAGTATGATCAGCTGGAACAGAAA	2220
QY	2070	TTCAGAAACCGAGCAGGATAAGGCAAACTTCTCTCTCCAGCAGCTCGAGGAAATCAAAACAC	2129
Db	2221	TTTAGAACTGAGCAGGATAAGGCTAACTTCTCTCTCAACAACTAGAGGAGATCAAGCAC	2280
QY	2130	CAAAATGCCCAAGCAAAAGCCATAGAGAAAAGGGAGGCGGTGAGCCAGGAAGCCGAACTG	2189
Db	2281	CAAAATTGCCAAGATAAAGCAATAGAGAAAGGAGGTGAGGTTGTGAGCCAGGAAGCTGAACTG	2340

Db 41114 CAGCTTAGACTCAAACCTGGTTGGCTTATCGCAAAGGATCGAGGAACCTGGAAGAGACCAAT 41055

Qy 1119 AAAAGCCTTCAGAAGGCGAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGG 1178

Db 41054 AAAAGTCTTCAGAAGGCGGAGGAAGAGCTTCAGGAGCTGAGAGATAAATTCCAAAGGG 40995

Qy 1179 GAATGTGAAAACCTCCAGTCTCATGGCGGAAGTGGAGAGTGTGCGCAAGCGCGTGTCTTGAG 1238

Db 40994 GAATGTGAAAACCTCCAGTCTCATGGCGGAAGTGGAAAAGTGTGCGCAAGCGCGTGTCTTGAG 40935

Qy 1239 ATGGAGGSCAAGGATGAAGAGATCACGAACACCGAGGCCCCAGTGCOCGGGAGCTGAAGAAG 1298

Db 40934 ATGGAGGSCAAGGATGAAGAGATCACTAAGACCGAGGCCCCAGTGCOCGGGAGCTGAAGAAG 40875

Qy 1299 AAGCTCCAAGAGGAAGAACACACAGCAAGGAACTTAGACTAGAAGTGGAGAAAGCTGCAG 1358

Db 40874 AAGCTCCAAGAGGAAGAACACACAGCAAGGAACTTAGACTAGAAGTGGAGAAAGCTGCAG 40815

Qy 1359 AAGAGGATGTCTGAGCTGGAGAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGC 1418

Db 40814 AAGAGGATGTCTGAACCTGGAGAAGCTGGAGGAGGCCCTTCAGCCGCAGCAAGTCGGAATGC 40755

Qy 1419 ACCCAGCTCCATCTGAACTCTGGAGAAGGAGAAGAACCTTAACCAAGACCTTGCTGAACGAG 1478

Db 40754 ACCCAGCTCCATTTGAACCTGGAGAAGGAGAAGAACCTTAACCAAGACCTTGCTGAACGAG 40695

Qy 1479 CTGGAGGTGGTCAAGAGTCAAGTGTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAG 1538

Db 40694 CTGGAGGTGGTCAAGAGTCAAGTGTAAAGAACTCGAGTGTCTGGAGAGCAGACTGGAGAAG 40635

Qy 1539 GCCGATTAAGCCTCAAAGATGACCTTACAAGCTGAAGTCTCTCACTGTGATGCTGGTG 1598

Db 40634 GTTGAGTTAAGCCTCAAAGACGACCTTCACGAAGCTGAAGTCTCTCACTGTGATGCTGGTG 40575

Qy 1599 GATGAGAGGAAAAATATGATGGAGAAAAATAAAGCAAGAACAGAGAGGAAAGTGCATGGGTTG 1658

Db 40574 GACGAGAGGAAAAAATCATGATGGAGAAAAATAAAGCAAGAACAGAGGAAAGTGGATGGGTTG 40515

Qy 1659 AATAAAAACTTTAAGGTGGAGCAGGGAAAAAGTCAATGATGTGACGGAAAAAGCTAATCGAG 1718

Db 40514 AATAAAAACTTCAAGGTGGAGCAGGGGAAAGTGATGGATGTGACGGAGAAAGCTAATCGAG 40455

Qy 1719 GAAAGCAAGAAGCTTTTAAAACTCAAATCTGAAATGGAGGAAAAAGGATACAGTCTGACA 1778

Db 40454 GAAAGCAAGAAGCTTTTAAAACTCAAGTCTGAAATGGAGGAAAAAGGTGTATAGTCTGACC 40395

Qy 1779 AAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGGAAGAAGAAAGGTCCTGTGAAC TG 1838

Db 40394 AAGGAGAGGGATGAGCTGATGGGTAAACTGAGAAGCGGAAGAAGAAAGGTCCTGTGAAC TG 40335

Qy 1839 AGCTGCAGTGTAGACTTAATAAGAAAGCGGTTTGATGGCATAGAGGAGGTAGAAAAGGAA 1898

Db 40334 AGCTGCAGCGTTGACTTGCTTAAAGAAAGCGGCTTGATGGCATAGAGGAGGTGAGAGGGAA 40275

Qy 1899 ATAAACCGAGGTAGTCTGCAAGGGGTCTGAGTTCACTGCCCCGGAAGACAATAAGATC 1958

Db 40274 ATCAACCGAGGAGGCTGTGCAAGGGGTCTGAGTTCACTGCCCCGGAAGACAATAAAATC 40215

Qy 1959 AGAGAACTAACGCTTGAAATCGAGAGACTGAAGAAAAACGGCTCCAGCAGTTGGAGGTGGTG 2018

Db 40214 AGGGAAC TGACGCTTGAAATCGAGAGACTGAAGAAAGCGGCTCCAGCAGCTGGAGGTGGTG 40155

Qy 2019 GAGGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAGCAGAAAGTTCAGAAAC 2078

Db 40154 GAGGGGACTTGATGAAGACCGAGGACGAATACGACCAGTTGGAGCAGAAATTCAGAAAC 40095

Qy 2079 GAGCAGGATAAGCGCAAACTTCTCTCCAGCAGCTCGAGGAAATCAAAACACCAAAATGGCC 2138

Db 40094 GAGCAGGATAAGCGCAAAATTTCTCTCTCCAGCAGCTCGAGGAAATTAAGCACCAAAATGGCC 40035

Qy 2139 AAGCACAAAGCCATAGAGAAAGGGGAGGCCGTGAGCCAGGAAGCCGAAC TGCGACACAGG 2198

Db 40034 AAGAACAAAGCCATAGAGAAAGGGGAGGCTGTGAGCCAGGAAGCCGAGCTGCGACACAGG 39975

Qy 2199 TTTCCGCTGAGGAGGCTAAAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAGGAG 2258

Db 39974 TTCCGGATGAGGAGGCTAAAAAGTCGTGACTTACAGGCTGAGGTGCAGGCTCTTAAAGAG 39915

Qy 2259 AAGATCCACGAGCTGATGAACAAAGGAAGACACAGCTGTCTCAGCTCCAAGTCGACTATTTCG 2318

Db 39914 AAGATCCACGAGCTGATGAACAAAGGAAGATCAGCTCTCTCAGCTCCAGTCCGACTATTTCG 39855

Qy 2319 GTCTTACGAAAGATTTATGGAAGAAAGAACTAAGAACAAAGAACATGGGGAGGAGGTC 2378

Db 39854 GTCTTACGAAAGGTTTATGGAAGAAAGAACTAAGAACAAAGAACATGGGGAGGAGGTC 39795

Qy 2379 CTCAATCTGACCAAGGAGCTAGAGCTTTCCNAAGCGCTACAGCCGAGCTCTCAGGCCGAGT 2438

Db 39794 CTCAACCTGACCAAGAGCTGGAGCTTTCNAAGCGCTACAGCCGAGCTCTCAGGCCGAGT 39735

Qy 2439 GGGAAACGGCCGAAGGATGTTGGACGTGCCCTGTGGCCCTCCACTGGGGTGCAGACCGAGGCG 2498

Db 39734 GGAATGGCCGAAGGATGTTGGACGTACCCGTGGCCTCCACTGGGGTGCAGACCGAGGCG 39675

Qy 2499 GTGTCCGGGATGCTCGGAGGAGGAGACCCCGGCTGTGTTTCAATTCGCAAAATCCTTCCAG 2558

Db 39674 GTGTCCGGGATGCTCGGAGGAGGAGACCCCGGCTGTGTTTCAATTCGCAAAATCCTTCCAG 39615

Qy 2559 GAGGAAAAATCACATCATGAGTAATCTTCGACAGGTAGGCCCTGAAGAAAAACCCATGGAACGG 2618

Db 39614 GAAGAAAAATCACATCATGAGTAATCTTCGACAGGTAGGCCCTGAAGAAAAACCCATGGAACGG 39555

Qy 2619 TCCTCGTCTCGACAGGTATCCCCCAGCAGCGAATGAGTCAACCATGAGGAAGTCTTGG 2678

Db 39554 TCCTCGTCTCGACAGGTATCCCCCAGCAGCCCAATGAACCTCACCATGAGGAAGTCTTGG 39495

Qy 2679 ATTCTTGGATGAGAAAAAGAGAAAAACGGTCTCTTCCACTCCGACGAGAGAAAAAGGCCCCAGG 2738

Db 39494 ATTCTTGGATGAGAAAAAGAGAAAAACGGACCTCAGCTCCCAGGAGAAAAAGGCCCCAGG 39435

Qy 2739 CCAAAACAGGTGCAGGGCAACCCCGGGAGTGGTCTTAGCAACAAAAGCAGGGCCAGCCCC 2798

Db 39434 CCAAAACAGGTACAGGGCAACCTTGGCGAGTGGTCTTAGCAACCAAAGCAGGGCCAGCCCC 39375

Qy 2799 CTACACATCCGTGTGACACAGATCATGAGAAACAGCACTGCCACCTTGGAGATCACAAAGC 2858

Db 39374 TTGCACATCCGTGTGACACCTGATCACGAGAAACAGCACTGCCACCTTGGAGATCACAAAGC 39315

Qy 2859 CCCACATCTGAAGAGTTTCTCTAGTACCAACCGTCATTCCTACCTTAGGCAACCCAGAAA 2918

Db 39314 CCCACATCTGAAGAGTTTCTCGAGTACCAACCGTCATTCCTACCTTAGGCAACCCAGAAA 39255

Qy 2919 CCAAGAAATAACCAATATTCCATCACCCCAATGTCAATGTGCGAAGGCTAGGCAACCCAGAAA 2978

Db 39254 CCACGGATCACCAATTTCCATCACCCCAATGTCAATGTGCGAAGGCTAGGCAACCCAGAAA 39195

Qy 2979 CCTACTCTCGGCCCAGAACGAGCCATGTCCCTGTACGATTACTACTATTTCAGAGAG 3038

Db 39194 CCTACTCTCGGCCCAGAACGAGCCATGTCCCTGTACGATTACTACTCAATTTCCAGAGAG 39135

Qy 3039 AAGAGCCCAGAAAGGTGGAAGGAGCGCCTTTGCGACAGGCCCTGCATCCCCCATCCAAAATC 3098

Db 39134 AAGAGCCCAGAAAGGTGGAAGGAGCGCGTTTGCAGACAGGCCCTGCATCCCCCATCCAGATA 39075

Qy 3099 ATGACGGTGTCAACATCTGCAAGTCCCACCTGAATCGTGTCTCTCTGAATCTCAGGAA 3158

Db 39074 ATGACGGTGTCAACCGTCTGACGGCCCCCTGAAATCGTGTCTCTCTGACTCTCAGGAA 39015

Qy 3159 GTGCCTATGGGAAGGACTATCCTCAAAGTCAACCCCGGAAAAACAAACTGTTCCAGCCCCC 3218

Db 39014 GTGCCTATGGGAAGGACTATCCTCAAAGTTACCCCGGAAAAACAAACTGTTCCAAACCCCC 38955

Qy 3219 GTGCGGAAGTACAACTCCAAATGCTAATATCATCATCACCGAAGCAATAAAAAATTCACATT 3278

Db 38954 ATGCGGAAGTACAAATGCCAATGCGAATATCATCACCGAAGCAATAAAAAATTCACATT 38895

QY 3279 CACCTGGGTTCTCAGTTTAAGCGATCTCCTGGGCCT---GCCGCTGAAGGCGTGAGCCCA 3335
|||||
Db 38894 CACCTGGGTTCTCAGTTTAAGCGATCTCCCGGCCTGCAGCAGCGGAAGCGTGAGTCCA 38835
|||||
QY 3336 GTTATCACCGTCCGGCCTGTCAACGTTGACAGCGGAGAGGAGGTTTCTACAGGCACAGTC 3395
|||||
Db 38834 GTTATCACCGTCCGGCCTGTCAACGTTAACAGCCGAGAGGAGGTTTCTACGGGCACAGTC 38775
|||||
QY 3396 CTTGCGTCTCCAGGAACCACTCTCTTCAAGACCCGGTGCTAGCAAAAGTGACCAAGCACT 3455
|||||
Db 38774 CTTGCGTCTCCAGGAACCACTCTCTTCAAGACCTGGTGCTAAACAAAGTGACCAAGCACT 38715
|||||
QY 3456 ATAACCTATAACCCCGGTCAACAACGTCTATCCACACGAGGAACCCCAATCAGTG 3506
|||||
Db 38714 ATAACCTATAACCCCGGTCAACAACGTCTATCCACACGAGGAACCCCAATCAGTG 38664
|||||

RESULT 13
AK027705
LOCUS AK027705 3559 bp mRNA linear PRI 01-AUG-2002
DEFINITION Homo sapiens cDNA FLJ14799 fis, clone NT2RP4001351, weakly similar
to Human ovarian cancer downregulated myosin heavy chain homolog
(Doc1) mRNA.
ACCESSION AK027705
VERSION AK027705.1 GI:14042583
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Ishibashi,T., Fujimori,K.,
Tanai,H., Kimata,M., Watanabe,M., Hiraoka,S., Ishii,S., Kawai,Y.,
Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K.,
Masuho,Y. and Kanehori,K.

TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3559)
AUTHORS Isogai,T. and Otsuki,T.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology; cDNA library construction;
5'- & 3'-end one pass sequencing and clone selection: Helix
Research Institute (supported by Japan Key Technology Center etc.)
and Department of Virology, Institute of Medical Science,
University of Tokyo.

FEATURES
source Location/Qualifiers
1..3559
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP4001351"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RP4"
/note="cloning vector: pME18SFL3~mRNA from NT2 neuronal
precursor cells after 2-weeks retinoic acid (RA)
induction."
64..2961
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB55310.1"
/db_xref="GI:14042584"
/translation="MLVDERQMHIQLGLQSQKVDLTQKLREEBEKLAITSKSKED
RQKLLKLEVDPEHKASRFQSQHEEMNAKLANQSHNRQLRLKLVGLTQRILEEETNK
NLQKAEELQLERDKIAKGCNSSLMAEVENLRKRVLEMEDKDEITKTESQCRELR
KKLQEEHHHSKELRLEVEKLOKRMSELEKLEEAFFSKSECTQLHLNLEKEKNLTKDL
LNELEVVKSRVKELECSERLEKAELSLKDDLTCLKSFTVLMVDERKNMMEKIQEER

CDS
64..2961
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB55310.1"
/db_xref="GI:14042584"
/translation="MLVDERQMHIQLGLQSQKVDLTQKLREEBEKLAITSKSKED
RQKLLKLEVDPEHKASRFQSQHEEMNAKLANQSHNRQLRLKLVGLTQRILEEETNK
NLQKAEELQLERDKIAKGCNSSLMAEVENLRKRVLEMEDKDEITKTESQCRELR
KKLQEEHHHSKELRLEVEKLOKRMSELEKLEEAFFSKSECTQLHLNLEKEKNLTKDL
LNELEVVKSRVKELECSERLEKAELSLKDDLTCLKSFTVLMVDERKNMMEKIQEER

ORIGIN

Query Match 56.0%; Score 2443.2; DB 9; Length 3559;
Best Local Similarity 81.6%; Pred. No. 0;
Matches 2918; Conservative 0; Mismatches 638; Indels 22; Gaps 7;
QY 753 GAAACCGCTAAGCGGCTCAACAAACTTCGAGATGAGCTTGTGAAGCTCAAGTCTCTCGCC 812
|||||
Db 1 GAAATGCTAAACGACTCAATAAACTAAGAGATGAGCTTGTAAACTCAAATCCTTTGCA 60
|||||
QY 813 CTCATGTTGGTGACGAGAGGCAGATGCACATCAGCAACTGGGCCTGCAGAGTCAGAAA 872
|||||
Db 61 CTCATGCTGGTGGATGAAAGACAAATGCACATTGCAACAACTTGGCCTGCAAAAGCCAGAAA 120
|||||
QY 873 GTCCAGGACCTCACTCAGAAAGCTGAGGGAGGAGGAAGAAAACTCAAAGCGGTCACTTAC 932
|||||
Db 121 GTACAGGATCTTACTCAGAAAGCTGAGGGAAGAAAGAGAAGCTCAAAGCCATTACTTCC 180
|||||
QY 933 AAATCCAAGGAAGACCGCCAGAAAGCTGCTCAAGTTAGAAGTGGAAGTTCGAACACACAGGCC 992
|||||
Db 181 AAATCCAAGGAAGACAGACAGAAATTGCTCAAGTTAGAAGTGGAAGTTCGAACACACAGGCT 240
|||||
QY 993 TCGAGGTTTTCCAGGAGCAGAAAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCAC 1052
|||||
Db 241 TCGAGGTTTTCTCAAGAGCATGAAGAGATGAACGCTAAACGTGGCTAATCAAGAGTCTCAC 300
|||||
QY 1053 AACCGGCAACTTCGACTCAAACTGGTTGGCTTATCGCAAAAGGATTGAGGAGCTGGAAGAG 1112
|||||
Db 301 AATAGGCAACTTTAGACTCAAGCTGGTTGGCTTTAAACCCAAAGAATCGAGGAGCTAGAAGAG 360
|||||
QY 1113 ACCAATAAAGCCTTCAGAAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAGAAAATTGCC 1172
|||||
Db 361 ACCAATAAAGCCTTCAGAAGGCAGAGGAAGAACTTCAAGAAATTAAGAGATATAAAATTGCC 420
|||||
QY 1173 AAAGGGAATGTGGAAGCTCCAGTCTCATGCGGGAAGTGGAGAGTCTCGCGCAAGCGCGTG 1232
|||||
Db 421 AAAGGGAATGTGGAAGCTCTAGCCTCATGCGGGAAGTGGAAAATCTTCGAAAGCGGTG 480
|||||
QY 1233 CTTGAGATGGAGGCAAGGATGAAGAGATCAAGAACCCGAGGCCAGTCCCGGAGCTG 1292
|||||
Db 481 CTTGAAATGGAAAGGTAAAGATGAGGAGATCACTAAACTCCAGTCCAGTAATTG 540
|||||
QY 1293 AAGAAGAAGCTCCAAGAGGAAGAACACCACAGCAAGGAACCTTAGACTAGAAAGTGAAGAG 1352
|||||
Db 541 AGGAAGAAGCTGCAAGAGGAAGAAACACCATAGTAAGGAGCTCAGACTTGAAGTTGAGAAG 600
|||||
QY 1353 CTGCAAGAAGAGATGTCTGAGCTGGAGAAGCTGGAGGAAGCGTTTACGCCGAGTAAAGTGC 1412
|||||
Db 601 CTACAGAAGAGAAATGTCTGAAGTGAAGAAATTGGAAGAAGCATTTAGCAAGAGTAAATCT 660
|||||
QY 1413 GAATGCACCCAGCTCCATCTGAACCTGGAGAAGGAGAAAGAACTTAACCAAGACCTGTG 1472
|||||
Db 661 GAGTGACCCAGCTACATTAAATCTGGAGAAGAAAGAACTTAACCAAGACCTGTGTA 720
|||||
QY 1473 AACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTG 1532
|||||
Db 721 AATGAATTGGAGGTGGTCAAGAGTCGAGTTAAAGAAATTTGGAATGTTCTGAAAGTAGATTG 780
|||||
QY 1533 GAGAAGGCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAAGTCTTCTCACTGTGATG 1592
|||||
Db 781 GAAAAGGCTGAATTAAAGCCTAAAAGATGATCTTACCAAGTTGAAGTCAATTACCGGTGATG 840
|||||

QY	1593	CTGGTGGATGAGAGGAAAAATATGATGGAGAAAAATAAAGCAAGAGAGAGGAAAGTGGAT	1652
Db	841	CTGGTTGATGAAGGAAAAATATGATGGAAAAATATAAAACAAGAGAGAGAAAGTGGAT	900
QY	1653	GGGTTGAATAAAACTTTAAGGTGGAGCAGGGAAAAAGTCATGGATGTGACCGGAAAAAGCTA	1712
Db	901	GGACTCAATAAAAAATTTAAGGTGGAACAAGGAAAAAGTTATGGATGTAACCTGAAAAACTA	960
QY	1713	ATCGAGGAAAGCAAGAAAGCTTTTAAAAACTCAAACTCTGAAATGGAGGAAAAAGGAGTACAGT	1772
Db	961	ATTGAAGAAAGTAAGAAACTTTTAAAAACTAAAACTGAAATCTGAAATGGAGGAAAAAGTATACAAC	1020
QY	1773	CTGACAAAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGAAGAAAGGTCTCTGT	1832
Db	1021	TTGACAAGAGAAAGAGATGAGTTGATAGGCAATTTGAAAGTGAAGAAAGAAAAATCCTCT	1080
QY	1833	GAACTGAGCTGCAGTGTAGACTTACTATAAGAAAGCGGCTTGATGGCATAGAGGAGTAGAA	1892
Db	1081	GAATTAAGCTGCAGTGTGACTTACTTAAAGAAAGACCTTGATGGTATAGAGGAAGTGAA	1140
QY	1893	AGGGAATAAACCGAGGTAGTGTGTCGAAGGGGTCTGAGTTACCTGCCCCGGAAGACAAT	1952
Db	1141	AGAGAAATAACAGAGGAAGGTACGAAAAAGGGTCTGAGCTCACCTGCCCCGGAAGATAAT	1200
QY	1953	AAGATCAGAGAACTAAACGCTTTGAAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAG	2012
Db	1201	AAGATTAAAGGAACTAACACTTTGAAATTTAGAGACTGAAGAAACGCTCTCCAACAATTGGAA	1260
QY	2013	GTGGTGGAGGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTTGGAGCAGAAAGTTC	2072
Db	1261	GTGGTCGAAGGGGATTTGATGAAGACAGAAAGTGTGATGATGATGATGATGATGATGATGAT	1320
QY	2073	AGAACCGAGCAGGATAAGGCCAAACTTCTCTCCAGCAGCTCGAGGAAATCAAAACACCAA	2132
Db	1321	AGAACTGAGCAGGATAAGGCTAACTTCTCTCTCAACAACCTAGAGGAGATCAAGCACCAA	1380
QY	2133	ATGGCCAAAGCACAAAGCCATAGAGAAAGGGGAGGCGGTGAGCCAGGAAGCGGAACCTGCGA	2192
Db	1381	ATTGCCAAGAATAAAGCAATAGAGAAAGGGTGAGGTTGTGAGCCAGGAAGCTGAACTGAGA	1440
QY	2193	CACAGGTTTCGGCTGGAGGAGCTAAAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTC	2252
Db	1441	CACAGATTTCGGTTGGAAGAGCTAAAAAGTCGAGACTTAAAGCCGGAGTACAAGCTCTT	1500
QY	2253	AAGGAGAAAGATCCACGAGCTGATGAACAAGGAAGACCGAGCTGTCTCAGCTCCAAGTCGAC	2312
Db	1501	AAAGAGAAGATTACGGAATTAATGAACAAGAAAGATCAGCTTTCTCAGCTCCAGGTAGAT	1560
QY	2313	TATTCGGTCTTCAGCAAGATTTATGGAAGAAGAAACTTAAGAAACAAGAACATGGGGAGG	2372
Db	1561	TATTCGTACTTCAACAAGATTTATGGAAGAAGAAATTAAGAAACAACAAACATGGGGCAG	1620
QY	2373	GAGGTCCTCAATCTGACCAAGGAGCTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGG	2432
Db	1621	GAGGTTCTCAATCTGACCAAGAGTTGGAGCTTTCCAAGCGCTACAGCAGAGCTCTTAGG	1680
QY	2433	CCGAGTGGAAACGGCCGAAGGATGGTGGACGTGCCCTGTGGCCTCCACTGGGGTGCAGACC	2492
Db	1681	CCCAGTGTAAATGGAATAAGAAATGGTGGATGTTCTCTGTGACGTCAACTGGAGTCCAAACT	1740
QY	2493	GAGCGGTGTGCGGGATGCTGCGGAGGAGAGACCCCGGCTGTGTTCAATTCGCAATCC	2552
Db	1741	GATGCAGTCAGCGTGAAGCAGCAGAGGAAGAAACGCCAGCTGTATTATATACGGAATCC	1800
QY	2553	TTCCAGGAGGAAAAATCACAATCATGAGTAATCTTCGACAGGTTAGCCCTGAAGAAACCCATG	2612
Db	1801	TTCCAGGAAGAAAAATCATATTATAGTAAATCTTCGGCAGGTGGGATTTGAAGAAACCCGTG	1860
QY	2613	GAACTGCTCGGTCTTCGACAGGTTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAG	2672
Db	1861	GAAAGATCTTCTGTTCTAGACAGGTTATCCTCCAGCAGCAAAATGAGCTCACTATGAAAG	1920

QY	2673	TCTTGGATTCTTGGATGAGAAAAAGAGAAAAACGGTCTTTCACCTCCGACGAGAAAAAGGG	2732
Db	1921	TCTTGGATTCCATGGATGAGAAAAGAGGAAAAACGGCCCTCCATCACTCAGGAGAAAGGG	1980
QY	2733	CCCAGGCCAAACCAGGTCAGGGCCACCCCGGGGAGCTGGTCCTAGCACCAAAAGCAGGGC	2792
Db	1981	CCCCGAACAAATTCAGTCCAGGGCACCCAGGAGAGGTAGTCTTTTCCACCAAGCAGGGC	2040
QY	2793	CAGCCCCCTACACATCCGTGTGACACCCAGATCATGAGAAACAGCACTGCCACCTGGAGATC	2852
Db	2041	CAGCCCCCTGCATATTCGAGTGACACCCAGACCAACGAGAACAGCACTGCGACTTTGGAGATA	2100
QY	2853	ACAAGCCCCACATCTGAAGAGTTTTTCTCTAGTACCACCGTCATTCCTACTAGGCAAC	2912
Db	2101	ACAAGCCCGACATCTGAAGAATTTTTTCTAGTACCACTGTCTATTCCTACTTAGGAAT	2160
QY	2913	CAGAAACCAAGAATAACCATTTATCCATCACCCCAATGTCTATGTCGCAAAAGCCCCAAAAGT	2972
Db	2161	CAGAAACCAAGAATAACCATTTATCCATCACCAACGTTATGCTCAAAAACAAAAAAGT	2220
QY	2973	GCAGATCCTACTCTCGGCCCAGAACGAGCCATGTCCCTGTCTCAGTATCTACTACTATTTCC	3032
Db	2221	GGAGATACTACTCTTGGCCAGAACGAGCCATGTCCCCAGTCACAATTACTACATTTTTC	2280
QY	3033	AGAGAGAGAGCCCCGGRAGGTGGAAGGAGCGCCTTTGCGACAGGCCCTGCATCCCCCATC	3092
Db	2281	AGAGAGAGAGACTCCAGAAAAGTGGAGAGGCGCATTTGCAGACAGGCCACATCCCCCTATT	2340
QY	3093	CAAAATCATGACGGTGTCAAACATCTGCAGTCCCACCTGAAATCGTGTCTCTCTGAATCT	3152
Db	2341	CAGATAATGACGGTGTCTACATCAGCAGCACCCAGCTGAGATTGCAGTTTCTCCGAATCC	2400
QY	3153	CAGGAAGTCCCTATGGGAAGGACTATCTCTCAAAGTCAACCCCGGAAAAACAAACTGTTCCA	3212
Db	2401	CAGGAATGCCCATGGGACGGACAACTCTCAAAGTCAACCCCGGAAAAACAGACTGTTCCA	2460
QY	3213	GCCCCGTGCGGAAGTACAACACTCCAATGCTAAATATCATCACCCAGCAAGAGCAATAAAAAT	3272
Db	2461	ACTCCAGTACGGAATACAACTCCAATGCCAATATCATAAACCAACAGAGGACAAATAAAAT	2520
QY	3273	CACATTCACTGGGTTCTCAGTTTAAGCGATCTCTTGGGCCCTGCCGTGAAGGCGTGAGC	3332
Db	2521	CACATTCACTTAGGGTCTCAGTTTAAACGGTCCCCTGGACTTCAGGTGAAGGAGTCAGT	2580
QY	3333	CCAGTTATCACCGTCCGGCTGTCAACGTGACAGCGGAGAAAGGAGTTTCTACAGGCACA	3392
Db	2581	CCAGTTATTACTGTCCGACCAAGTAAACGTGACAGCGCAAAAGGAGGTTTCCAACCGCACT	2640
QY	3393	GTCCTTCGCTCTCCAGGAACCACTCTCTTCAAGACCCGGTGTAGCAAAAGTGACCAGC	3452
Db	2641	GTCCTTCGCTCTCCAGGAATCACTCTCTCTCACGGCCCTGGTGCAAGCAAAAGTGACGAGC	2700
QY	3453	ACTATAACTATAACCCCGGTCAACAGTCATCCACACGAGGAACCCAATCAGTGTCAAGGA	3512
Db	2701	ACTATCACCATAAACACCGGTCAACACGTCTCTGTCTCGAGGAACCCAGTCAGTGTCAAGGA	2760
QY	3513	CAAGATGGGTCACTCTCAGCGGCCCTACCCCAACCCGCATTCCTATGTCAAAAAGGTATGAAA	3572
Db	2761	CAAGACGGGTCACTCCAGCGGCCCTACACCCACCCGCATTCCTATGTCAAAAAGGTATGAAA	2820
QY	3573	GCTGGAAGCCAGTAGTGGCAGCTCAGGAGCAGGAAATCTGACCAAAATTCAGCCTCGA	3632
Db	2821	GCAGGAAGCCAGTAGTGGCAGCCCCCAGGAGCAGGAAATCTGACCAAAATTCGAGCCTCGA	2880
QY	3633	GCTGAGACTCAGTCTATGAAAAATAGAGCTGAAGAAATCTGCAGCCAGCAGCACTGCCCTCT	3692
Db	2881	GCTGAGACTCAGTCTATGAAAAATAGAGCTGAAGAAATCTGCAGCCAGCAGCACCCCTCT	2940
QY	3693	CTTGAGGGGGGAAGGCTGAGGCGAGTGGCTAAGGGGGTATGTTGTAAGGATGCTACTG	3752
Db	2941	CTCGAGGGGGGAAGGCTGAGGCGAGTGGCTAAGGGGGTATGTTGTGCAGATGCTACTG	3000
QY	3753	CTGCAGTGGAAAAACAAACCTTCTCTGTGTGCCAACCCCTTTCTCTGTA-CTACTAATTTAAGT	3811

Db 3234 CATCTGAGC-----CTGTTTCCTCATCAGTTAGATGTGGGACTCAAT--CACA 3280

QY 4052 TGCTCCCAAAGGCTGGCTCCCAATTTTCCTAATTGTAAGCCAACTTTAATAGACTCAGTT 4111

Db 3281 CGCTCTTCAAGTCCGGCTCCCATATTTCCTAATTGCAAGCCAAATTTAATGTACCTTGTT 3340

QY 4112 CTGTGAT-TTTTTTTTCCAAAAAAAATAATTTTGAATAGGACAGAGTTTAAACAGTTGT 4170

Db 3341 CCACAATAATTTTTATTAAAAAATCCTATTACAAAAAAGACATACCTTTAACTATTGT 3400

QY 4171 CATTTGCACTATCAAGCCATGAGTTTGATATATATGGGTTATAAGAAA-AGAATACTTTCA 4229

Db 3401 CATTTGCCCTCTTTCACATCATGAATTTGCTTTATGTGCTGGA AAAACATCACA TAGCTA 3460

QY 4230 GAGCTATCACAGGCTCTTAAGCTTTTGGAAAAACAAAAGCCCTAATATGACCTCAGGA 4289

Db 3461 TCACAGGGCCTGGACCTCTAAAAATTTTGCAAAAAACAAAAGGTTCTAAGATGATTTTCAGGA 3520

QY 4290 AACAATTTGAACATGAATAAAAAATGGAATGAACCTGTG 4327

Db 3521 AATAATGTGAACATGTATAAAAAATGGAATGAAATATG 3558

Search completed: September 8, 2004, 06:31:55
Job time : 26590 secs

THIS PAGE LEFT (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 7, 2004, 21:24:04 ; Search time 323 Seconds
(without alignments)
7497.850 Million cell updates/sec

Title: US-10-788-793-1
Perfect score: 4364
Sequence: 1 ccactgggtttcttcaaggga.....aaaaaaaaaaaaaaaaaaaaa 4364

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3025.6	69.3	4892	4	US-09-620-312D-1065
2	130	3.0	418	4	US-09-621-976-10456
3	99.6	2.3	7218	1	US-08-232-463-14
4	93.6	2.1	3489	2	US-08-728-323A-1
5	93.6	2.1	3489	4	US-09-298-568-1
6	93.6	2.1	3489	4	US-09-410-399-1
7	93.6	2.1	32207	2	US-08-770-379-20
8	93.6	2.1	32207	3	US-08-757-669A-20
9	93.6	2.1	32207	4	US-09-230-371A-20
10	70.4	1.6	5919	4	US-08-875-435B-2
11	65.6	1.5	1926	4	US-09-249-585A-2
12	65.6	1.5	1926	4	US-09-410-399-3
13	65.6	1.5	2580	3	US-09-050-863-2
14	65.6	1.5	2580	4	US-09-359-081-2
15	65.6	1.5	5452	2	US-09-130-114-1
16	65.6	1.5	8705	4	US-09-647-344A-14
17	65.6	1.5	9600	3	US-08-910-647-1
18	65.6	1.5	9600	4	US-09-620-925-1
19	65.6	1.5	10596	1	US-07-884-811-15
20	65.6	1.5	10596	1	US-07-885-971-15
21	65.6	1.5	10596	1	US-08-087-783A-15
22	65.6	1.5	10596	1	US-08-194-088B-15
23	65.6	1.5	10596	2	US-08-194-087-15
24	65.6	1.5	10596	5	PCT-US93-04648-15
25	65.6	1.5	16080	4	US-09-724-566A-48
26	65.4	1.5	4722	4	US-08-979-608A-14
27	65.4	1.5	4722	4	US-09-517-849-14

28	65.4	1.5	4722	4	US-09-616-289-14	Sequence 14, Appl
29	64.8	1.5	6644	4	US-08-875-435B-5	Sequence 5, Appli
30	60.2	1.4	2338	1	US-08-425-069-1	Sequence 1, Appli
31	60.2	1.4	2338	2	US-08-317-844B-1	Sequence 1, Appli
32	59.2	1.4	16442	3	US-08-781-891-208	Sequence 208, App
33	59.2	1.4	16442	4	US-09-618-166-208	Sequence 208, App
34	55.2	1.3	51259	3	US-08-781-891-209	Sequence 209, App
35	55.2	1.3	51259	4	US-09-618-166-209	Sequence 209, App
36	54.6	1.3	1638	4	US-09-616-289-46	Sequence 46, Appl
37	54.6	1.3	4697	4	US-08-979-608A-17	Sequence 17, Appl
38	54.6	1.3	4697	4	US-09-517-849-17	Sequence 17, Appl
39	54.6	1.3	4697	4	US-09-616-289-17	Sequence 17, Appl
40	54.2	1.2	340	1	US-08-182-175A-104	Sequence 104, App
41	54.2	1.2	340	5	PCT-US92-06412-104	Sequence 104, App
42	54.2	1.2	1588	4	US-09-490-291-7	Sequence 7, Appli
43	53	1.2	2887	2	US-08-533-306A-3	Sequence 3, Appli
44	53	1.2	2887	2	US-08-742-923A-3	Sequence 3, Appli
45	52	1.2	243	1	US-08-182-175A-56	Sequence 56, Appl

ALIGNMENTS

RESULT 1
US-09-620-312D-1065
; Sequence 1065, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 1065
; LENGTH: 4892
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (634)..(4275)
US-09-620-312D-1065

Query Match 69.3%; Score 3025.6; DB 4; Length 4892;
Best Local Similarity 82.5%; Pred. No. 0;
Matches 3593; Conservative 0; Mismatches 734; Indels 27; Gaps 10;

QY 1 CCAC TGGGTTCTTCAAGGATAAACCGCGC-GAAAGAACACACCATTTGGTT-AAGGAG 58
|||||
Db 558 CCAC TGGGTTCTTCAAGGATAAACTACCTACATAGGACATACCTTTGGTTAAAGGAG 617

Db 2778 TAAAAGTCGAGACTTAAAAGCGGAAGTACAAGCTCTTTAAAGAGAAGATTTCACGAATTAAT 2837

QY 2276 GAACAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTATTTCGGTCCCTTCAGCGAAAAGATT 2335

Db 2838 GAACAAGAAGATCAGCTTCTCAGCTCCAGGTAGATTATTCTGTACTTCAACAAGAATT 2897

QY 2336 TATGGAAGAAGAAACTAAGAACAAGAACATGGGGAGGAGGTCTCAATCTGACCAAGGA 2395

Db 2898 TATGGAAGAAGAAATAAGAACAANAACATGGGGCAGGAGGTCTCAATCTGACCAAGA 2957

QY 2396 GCTAGAGCTTTCCAAGCGTACAGCCGAGCTCTCAGGCGAGTGGGAACGGCCGAAGGAT 2455

Db 2958 GTTGGAGCTTTCCAAGCGCTACAGCAGAGCTCTTAGGCCAGTGTGAATGGAAGAAGAAAT 3017

QY 2456 GGTGGACGTGCCGTGCGCTCCACTGGGTCAGACCGGAGCGGTGTGCGGGATGCTGC 2515

Db 3018 GGTGGATGTTCTGTGACGTCAACTGGAGTCCAAACTGATGCAGTCAGCGGTGAAGCAGC 3077

QY 2516 GGAGGAGGAGACCCCGCTGTGTTCATTTCGCAAAATCCTTCCAGGAGGAAATCACATCAT 2575

Db 3078 AGAGGAAGAAACGCCAGCTGTATTTCATACGGAATCCTTCCAGGAAGAAATCATATTAT 3137

QY 2576 GAGTAATCTTCGACAGGTAGGCTTGAAGAAACCCATGGAACGGTCTCTGGTCTCTCGACAG 2635

Db 3138 GAGTAATCTTCGGCAGGTGGATTGAAGAAACCCGTGGAAGAGATCTTCTGTCTAGACAG 3197

QY 2636 GTATCCCCAGCAGCGAATGAGCTCACCATGAGGAATCTTGGATTCTTGGATGAGAAA 2695

Db 3198 GTATCCTCCAGCAGCAATGAGCTCACTATGAGAAAGTCTTGGATTCCATGGATGAGAAA 3257

QY 2696 AAGAGAAAACGGTCTTCCACTCCGAGGAGAAAGSGCCAGGCCAAACCCAGSGTGCAGG 2755

Db 3258 GAGGGAACACGGCCCTCCATCACTCAGGAGAAAGSGCCCGAACAAATTCACGTCCAGG 3317

QY 2756 GCACCCGGGAGCTGCTTAGCACCAAGCAGGSGCCAGCCCCCTACACATCCGTGTGAC 2815

Db 3318 GCACCCAGGAGGTAGTCTTTACCAAGCAGGSGCCAGCCCCCTGCATATTTCAGTGAC 3377

QY 2816 ACCAGATCATGAGAACAGCACTGCCACCCCTGGAGATCACAAAGCCCCCACATCTGAAGAGTT 2875

Db 3378 ACCAGACCACGAGAACAGCACTGCGACTTTGGAGATAACAAGCCCCGACATCTGAAGAAAT 3437

QY 2876 TTTCTCTAGTACCACCGTCAATTCCTTACCTTAGGCAACCAAGAAACCAAGAAATACCATTA 2935

Db 3438 TTTTCTAGTACCACCTGTCAATCTTACCTTAGGGAATCAGAAACCAAGAAATACCATTA 3497

QY 2936 TCCATCACCCAATGTATGTGCGAAAGCCCCAAAGTCAGATCCTACTCTCGGCCCAGA 2995

Db 3498 TCCATCACCAACCGTTATGCCTCAAAAACAANAAGTGGAGATACTACTCTTGGCCCCAGA 3557

QY 2996 ACGAGCCATGTCCCTCTCAGGATTACTACTATTTCAGAGAGAAGAGCCCCGGAAGGTGG 3055

Db 3558 ACGAGCCATGTCCCCAGTCACAATTACTACATTTTCCAGAGAGAAGACTCCAGAAAAGTGG 3617

QY 3056 AAGGAGCGCCTTTGCGACAGGCGCTGCATCCCCCATCCAAATCATGACGGTGTCAACATC 3115

Db 3618 AAGAGGCGCATTTGCAGACAGGCCCCACATCCCCTATTTCAGATAATTGACGGTGTCTACATC 3677

QY 3116 TGCAGCTCCCACGTGAATCGCTGTCTCTCTGAAATCTCAGGAAGTGCCTATGGGAAGGAC 3175

Db 3678 AGCAGCACCACTGAGATTGCAGTTTCTCCCGAATCCCAGGAATGCCCATGGGACGGAC 3737

QY 3176 TATCCTCAAAGTCACCCCGGAAAAACAACACTGTTCCAGCCCCCGTGCAGGAAGTACAACTC 3235

Db 3738 AATCCTCAAAGTCACCCCGGAAAAACAGACTGTTCCAACTCCAGTACGGAATAACAACCTC 3797

QY 3236 CAATGCTAATATCATCACCGAAGACAAATAAATTCACATTTCACCTGGGTCTCAGTT 3295

Db 3798 CAATGCCAATATCATTAACACAGAGGACAAATAAATTCACATTTCACCTTAGGGTCTCAGTT 3857

QY 3296 TAAGCGATCTCTGGCCTGCCGCTGAAGCGGTGAGCCCCAGTTATCACCCGTCGGCCTGT 3355

Db 3858 TAAACGGTCCCCTGGACTTCAGGTGAAGGAGTCACTCCAGTTATTACTGTCCGACCACT 3917

QY 3356 CAACGTGACAGCGGAGAGGAGGTTTCTACAGGCACAGTCTTGGCTCTCCAGGAACCA 3415

Db 3918 AAACGTGACAGCGGAAAAAGGAGGTTTCCACCGGCACTGTCTTGGCTCTCCAGGAATCA 3977

QY 3416 CCTCTCTTCAAGACCCCGTGTCTAGCAAAAGTGACCAGCACTATAACTATAACCCCGGTAC 3475

Db 3978 CCTCTCTCAGCGCTGTGCAAGCAAAAGTGACGAGCACTATCACCAATAACCCCGGTAC 4037

QY 3476 AACGTTCATCCACACAGGAACCCCAATCAGTGTCAAGCAAGATGGGTCTCTCAGCGGCC 3535

Db 4038 AACGTTCATCTGCTCGAGGAACCCAGTCAGTGTCAAGCAAGACGGGTCTATCCAGCGGCC 4097

QY 3536 TACCCCCACCCGATTCCTATGTCAAAAAGGTATGAAAAGCTGGAAGCCAGTAGTGGCAGC 3595

Db 4098 TACACCCACCCGATTCCTATGTCAAAAAGGTATGAAAAGCAGGAAGCCAGTAGTGGCAGC 4157

QY 3596 CTCAGGAGCAGGAAATCTGACCAAAATTCAGCCTCGAGCTGAGACTCAGTCTATGAAAAT 3655

Db 4158 CCCAGGAGCAGGAAATCTGACCAAAATTCGAGCCTCGAGCTGAGACTCAGTCTATGAAAAT 4217

QY 3656 AGAGCTGAAGAAATCTGCAGCCAGCAGCACTGCCCTCTTTCGAGGGGGGAAGGGCTGAGG 3715

Db 4218 AGAGCTGAAGAAATCTGCAGCCAGCAGCACCCACTCTCTCGAGGGGGGAAGGGCTGAGG 4277

QY 3716 GCAGTGGCTAAGGGGTATGTTGTAAGGATGCTACTGCTGCAGTGGAAAACAAACCTTCCT 3775

Db 4278 GCAGTGGCTAAGGGGTATGTTGTCAGATGCTACTGCTGCCGTGAAAGTGAACTTCAT 4337

QY 3776 CTGTGCCAACCCCTTTCCTTGTA - CTACTAATTTAAGTTTTAAATATCTTGTGTTATAAAAT 3834

Db 4338 CTGTTTGTGCCAGTTCTTTACATGTACTAATTTAAGTTTTAAATATTTGTTTATAAAAT 4397

QY 3835 AACCATTTAATAGCCATGCACCCCTCCCTCAATTTTGTGTCATCTGTTTCAATGCGGGAA 3894

Db 4398 AACCAACTAATAACCAATTTGTCTT - TCCCAATTTGTGTCATTTGTTTGTGCTGGGAA 4455

QY 3895 TAGAATTAATTAGCAGAAATTTCTGTTTGTGTAAGTGTCTGTTGAAGATGTTGGTCCAGTT 3954

Db 4456 CA - - - AAATTAGCAAAACTATTGCTTGTCTGCTAGAACCCAGGGCGTGGTTCTAGTTC 4511

QY 3955 CAGTTTACTTCTAGCATGTGGCCCCCATTTCAAGGTAGCTCACGAGTTGTGAAGCCCTCAA 4014

Db 4512 CAGTTTGTCTTACCAAGTGGACCCCATCAATAGACCCATCTGAG - - - - - CCTG 4560

QY 4015 TATCGTCACCGGAGAGATTTGAGGACCACATATACATATGCTCCCAAGGCTGGCTCCCAA 4074

Db 4561 TTTCTCATCAGTTAGATGTGGGACTCAAT - - CACACGCTCTTCAAGTCCGGCTCCCAT 4618

QY 4075 TTTTCTTAATTGTAAGCCAACTTTAATAGACTCAGTTCTGTGAT - TTTTTCCTCAAAAA 4133

Db 4619 ATTTCTTAATTGCAAGCCAAATTTAATGTACCTGTTCCACAATNAATTTTTTATTAAAAA 4678

QY 4134 AAAAAATTTTGAATAGGACAGAGTTTAAACAGTTGTCTATTTTGCACATATCAAGCCATGA 4193

Db 4679 AATCCTATTACAAAATAAGACATACTTTAACTATTGTCTATTTGCCCTCTTTCACATCATGA 4738

QY 4194 GTTTGATATATGGGTATAAGAAA - AGAATACTTTTCAGAGCTATCACAGGCTCTCTAAAC 4252

Db 4739 ATTTGCTTTATGTCTGGAAAAAACATCACATAGCTATCACAGGGCTGGACCTCTAAAA 4798

QY 4253 TTTTGGAAAAACAAAAGCCCTAATATGACCTCAGGAAACAAATTTGAACATGAAATAAAA 4312

Db 4799 TTTTGCAAAAACAAAAGGTTCTAAGATGATTTTCAGGAAATAATGTGAACATGTAATAAAA 4858

QY 4313 TGGAAATGAACGTGGAATCTTTAAAAAATAAAAAA 4346

Db 4859 TGGAAATGAATATGGAATCCTTNAAAAAAATAAAAAA 4892

```
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 10456
; LENGTH: 418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-10456

Query Match      3.0%; Score 130; DB 4; Length 418;
Best Local Similarity 79.4%; Pred. No. 1.2e-22;
Matches 154; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 307 ACCTCATCCAGCTCCTGAGTATCATGGAAGGGAGTTGCGAGGCTCGAGAGATGTATCC 366
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 331 ACATTGTCAGTCTGATCTGTCTGCTGCACCTTCTATTACAGCCAGAGAGATGTATCC 272
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 367 ACATGCTGAGACAGAGAAAACCAAGCCCGAGGTTCTGGAGGCACACTATGGATCTGCAG 426
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 271 ACATGCTGAACAGACAGAGAAAACCAAGCCTGAGGTTCTGGAGGCTCATTACGGGCTGCGG 212
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 427 AACCTGAGAAAGTGTCTCGGTCCTGCACCGAGATGCCATCCTTGCTCAAGAGAGTCCA 486
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 487 TAGGAGAGACGTC 500
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 151 TAGGAGAGATGTC 138
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
US-08-232-463-14/c
; Sequence 14, Application US/082322463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-F1s
US-08-232-463-14

Query Match      2.3%; Score 99.6; DB 1; Length 7218;
Best Local Similarity 1.6%; Pred. No. 1.9e-14;
Matches 6; Conservative 264; Mismatches 108; Indels 0; Gaps 0;

QY 1595 GGTGGATCAGAGGAAAATATGATGGAGAAAATAAAGCAAGAGAGAGGAAAGTGGATGG 1654
Db ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 1655 GTTGAATAAAACTTTAAGGTGGAGCAGGGGAAAAGTCATGGATGTGACGGAAGAACTAAT 1714
Db ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 1374 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1315
QY 1715 CGAGGAAAGCAAGAGCTTTTAAAACTCAAAATCTGAAATGGAGGAAAGGAGTACAGTCT 1774
Db ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 1775 GACAAAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGAAGAAAGGTCTGTGA 1834
Db ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 1254 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1195
QY 1835 ACTGAGCTGCAGTGTAGACTTACTAAAGAGCGGCTTGATGGCATAGAGGAGGTAGAAAG 1894
Db ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 1895 GGAATAAAACCGAGTGTGTCGCAAGGGGTCTGAGTTCACTGCCCGGAAGACAATAA 1954
Db ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 1955 GATCAGAGAACTAACGCT 1972
Db ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
US-08-728-323A-1
; Sequence 1, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
;
US-08-728-323A-1

```

Query Match 2.1%; Score 93.6; DB 2; Length 3489;
Best Local Similarity 44.6%; Pred. NO. 4e-13;
Matches 415; Conservative 0; Mismatches 509; Indels 6; Gaps 1;

QY	519	GAGCTGGACAGACTGGAGGAAAAAGCAGAAAGGAGACGCTACCGCCGCATGTCTAGAGCAGCTG	578
Db	1717	GAGCCACAGCAGCGGGAGCCACAGCAGCGGGAGGCCACAGCAGCGGGAGCCACAGCAGCGG	1776
QY	579	CTGCTGGCTGAGAAGTGTACACAGGCGCACCGTGTACGAGCTGGAGAACGAGAAAGCACAAG	638
Db	1777	GAGCCACAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAG	1836
QY	639	CACACTGACTACATGAACAAAGAGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGAGAG	698
Db	1837	GATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAG	1896
QY	699	AGGTTGAAAAAGCTCTTGAACAAGAAAAAGCTTACCAAGCCCGCAAGAAAAAGGAAAAAC	758
Db	1897	GATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAG	1956
QY	759	GCTAAGCGGCTCAACAAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCATG	818
Db	1957	GATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAG-----GATGAGCAG	2010
QY	819	TTGGTGGACGAGAGGCAGATGCACATCGAGCAACTGGGCCCTGCAGAGTCAGAAAGTCCAG	878
Db	2011	CAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAG	2070
QY	879	GACCTCACTCAGAAGCTGAGGGAGGAGGAGAAAAAACTCAAAGCGGTCACTTACAAATCC	938
Db	2071	CAGCAGGATGAGCAGCAGCAGGATGAGCAGGAGCAGCAGGATGAGCAGGAGCAGCAGGAT	2130
QY	939	AAGGAAGACCGCCAGAAGCTGCTCAAGTTAGAGTGGACTTCGAAACAAGGCCCTCGAGG	998
Db	2131	GAGCAGGAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAG	2190
QY	999	TTTTTCCAGGAGCAGAGAGATGAACGCCAAATTGGCGAATCAAGAATCTCACAACCCGG	1058
Db	2191	CAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAG	2250
QY	1059	CAACTTCGACTCAAACTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGAAGAGACCAAT	1118
Db	2251	GATGAACAGGAGCAGGAGGAGCAGGAGGAGCAGCAGGAGGAGCAGGAGCAGGAGTATAGAG	2310
QY	1119	AAAAGCCTTCAGAAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGG	1178
Db	2311	GAGCAGGAGCAGGAGTTAGAGGATCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTA	2370
QY	1179	GAATGTGGAATCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTGTGAG	1238

QY 939 AAGGAAGACCGCAGAAAGCTGCTCAAGTTAGAAAGTGGAGCTTCGAAACACAAGGCTCGAGG 998
Db 2131 GAGCAGGAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAG 2190
QY 999 TTTTCCAGGAGCAGCAGGAGATGAACGCCAAATTTGGCGAATCAAGAACTCTCAACACCGG 1058
Db 2191 CAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAG 2250
QY 1059 CAACTTCGACTCAAACTGGTTGGCTTATCGCAAGGATTTGAGAGCTGGAAGACCAAT 1118
Db 2251 GATGAACAGGAGCAGCAGGAGCAGGAGCAGGAGCAGGAGCAGGAGCAGGAGTTAGAG 2310
QY 1119 AAAAGCCTTCAGAAAGCAGAGGAGAGCTCCAGGAGCTGAGAGAGAAAATTTGCCAAAGG 1178
Db 2311 GAGCAGGAGCAGGAGTTAGAGGATCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTA 2370
QY 1179 GAAATGTGMAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTCGGCAAGCGCTGCTTGAG 1238
Db 2371 GAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAG 2430
QY 1239 ATGGAGGCAAGGATGAAGAGATCACGAAGACCGAGGCCAGTCCCGGAGCTGAAGAAG 1298
Db 2431 TTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAG 2490
QY 1299 AAGCTCCAGAGGAAGAACACCAACAGCAAGGAACTTAGACTAGAAGTGGAGAAAGCTGCAG 1358
Db 2491 GAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGTGGAGAGCAAGAGCAGGAG 2550
QY 1359 AAGAGGATGTCTGAGCTGGAGAAAGCTGGAGGAAGCGTTTCAGCCGGAGTTAAGTCGGAATGC 1418
Db 2551 GTGGAAGAGCAAGAGCAGGAGCAGGAGCAAGAGCAAGAGCAAGAGCAAGAG 2610
QY 1419 ACCCAGCTCCATCTGAACCTGGAGAAAGGAG 1448
Db 2611 CAGGAGCAGGAGGAGCAGGAGGAGCAGGAG 2640

RESULT 6
US-09-410-399-1
; Sequence 1, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; APPLICANT: Cotter, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; FILE REFERENCE: UM-03778
; CURRENT APPLICATION NUMBER: US/09/410,399
; CURRENT FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-1

Query Match 2.1%; Score 93.6; DB 4; Length 3489;
Best Local Similarity 44.6%; Pred. No. 4e-13;
Matches 415; Conservative 0; Mismatches 509; Indels 6; Gaps 1;
QY 519 GAGCTGGACAGACTGGAGGAAAAGCAGAGGAGACGTACCGCGCATGCTAGAGCAGCTG 578
Db 1717 GAGCCACAGCAGCGGGAGCCACAGCAGCGGGAGCCACAGCAGCGGGAGCCACAGCAGCGG 1776
QY 579 CTGCTGGCTGAGAAGTGTACAGGCGCACCGTGTACAGCTGGAGACGAGAACGACAAG 638
Db 1777 GAGCCACAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAG 1836
QY 639 CACACTGACTACATGAACAAGAGCGGACGACTTCACCAACCTGCTGGAGCAGGAGCGAGAG 698
Db 1837 GATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAG 1896

QY 699 AGTTTGAAGAGCTCCTTGAACAAAGAAAAGCTTACCAAGCCCGCAAGAAAAGGAAAAC 758
Db 1897 GATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGCAG 1956
QY 759 GCTAAGCGGCTCAACAACTTCGAGATGAGCTTGTGAAGTCAAGTCTTCGCCCTCATG 818
Db 1957 GATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGCAGCAGCAG 2010
QY 819 TTGGTGGACGAGAGGCGAGATGCACATCGAGCAACTGGSCCTGCGAGATCAGAAAAGTCCAG 878
Db 2011 CAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAG 2070
QY 879 GACCTCACTCAGAAAGCTGAGGGAGGAGGAAGAAAACCTCAAAGCGGTCACTTACAAATCC 938
Db 2071 CAGCAGGATGAGCAGCAGCAGGATGAGCAGGAGCAGCAGGATGAGCAGGAGCAGCAGGAT 2130
QY 939 AAGGAAGACCGCCAGAAAGCTCAAGTTAGAAAGTGGACTTCGAAACACAAGGCCTCGAGG 998
Db 2131 GAGCAGGAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAG 2190
QY 999 TTTTCCAGGAGCAGCAGGATGAACCGCAAAATTTGGCGAATCAAGAACTCTCAACACCGG 1058
Db 2191 CAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGCAG 2250
QY 1059 CAACTTCGACTCAAACTGGTTGGCTTATCGCAAGGATTTAGAGGAGCTGGAGAGACCAAT 1118
Db 2251 GATGAACAGGAGCAGCAGGAGGAGCAGGAGCAGCAGGAGGAGCAGGAGCAGGAGTTAGAG 2310
QY 1119 AAAAGCCTTCAGAAAGCAGAGGAAAGAGCTCCAGGAGCTGAGAGAGAAAATTTGCCAAAGG 1178
Db 2311 GAGCAGGAGCAGGAGTTAGAGGATCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTA 2370
QY 1179 GAATGTGMAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTCGCAAGCGCTGCTTGAG 1238
Db 2371 GAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAG 2430
QY 1239 ATGGAGGCAAGGATGAAGAGATCACGAAGACCGAGGCCAGTCCCGGAGCTGAAGAAG 1298
Db 2431 TTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAG 2490
QY 1299 AAGCTCCAGAGGAAGAACACCAACAGCAAGGAACTTAGACTAGAAGTGGAGAAAGCTGCAG 1358
Db 2491 GAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAG 2550
QY 1359 AAGAGGATGTCTGAGCTGGAGAAAGCTGGAGGAAGCGTTTCAGCCGGAGTTAAGTCGGAATGC 1418
Db 2551 GTGGAAGAGCAAGAGCAGGAGCAGGAGCAAGAGCAAGAGCAAGAGCAAGAG 2610
QY 1419 ACCCAGCTCCATCTGAACCTGGAGAAAGGAG 1448
Db 2611 CAGGAGCAGGAGGAGCAGGAGGAGCAGGAG 2640

RESULT 7
US-08-770-379-20/c
; Sequence 20, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.

```

; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-770-379-20
,
Query Match          2.1%; Score 93.6; DB 2; Length 32207;
Best Local Similarity 44.6%; Pred. No. 1.3e-12;
Matches 415; Conservative 0; Mismatches 509; Indels 6; Gaps 1;

QY 519 GAGCTGGACAGACTGGAGGAAACGAGAAGGAGACGTACCGCCGCATGCTAGAGCAGCTG 578
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 20280 GAGCCACAGACGCGGGAGCCACAGCAGCGGGAGCCACAGCAGCGGGAGCCACAGCAGCGG 20221

QY 579 CTGCTGGCTGAGAAGTGTACAGAGGCGCACCGTGTACGAGCTGGAGAACGAGAAGCAACAG 638
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 20220 GAGCCACAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAG 20161

QY 639 CACACTGACTACATGAACAAGAGCGAGCAGACTTCACCAACTGTGTGAGCAGGAGCGAGAG 698
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 20160 GATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAG 20101

QY 699 AGGTTGAAAAAGCTCCTTGAACAAGAAAAAGCTTACCAAGCCGCAAGAAAAAGGAAAAAC 758
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 20100 GATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAG 20041

QY 759 GCTAAGCGGCTCAACAAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCATG 818
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 20040 GATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGCAGCAGCAGCAG 19987

QY 819 TTGGTGGACGAGAGGCAGATGCACATCGAGCAACTGGGCCCTGCAGAGTCAGAAAGTCCAG 878
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 19986 CAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAG 19927

QY 879 GACCTCACTCAGAAAGCTGAGGGAGGAGGAAGAAAACTCAAAGCGGTCACTTACAAATCC 938
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 19926 CAGCAGGATGAGCAGCAGCAGGATGAGCAGGAGCAGCAGGATGAGCAGGAGCAGCAGGAT 19867

QY 939 AAGGAAGACCGCCAGAAGCTGCTCAAGTTAGAAGTGGACTTCGAAACACAAGGCCCTCGAGG 998
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 19866 GAGCAGGAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAG 19807

QY 999 TTTTCCAGGACGACGAAGAGATGAACGCCAAATTTGGCGAATCAAGAAATCTCACAAACCGG 1058
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 19806 CAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAG 19747

QY 1059 CAACTTCGACTCAAACTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGAAGAGACCAAT 1118
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 19746 GATGAACAGGAGCAGCAGGAGGAGCAGGAGCAGCAGGAGGAGCAGGAGCAGGATTAGAG 19687

QY 1119 AAAAGCCTTCAGAAGGCAGAGGAGAGGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGG 1178
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 19686 GAGCAGGAGCAGGAGTTAGAGGATCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGGTTA 19627

```

QY	1179	GAATGTGGAAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTGTGAG	1238
Db	19626	GAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAG	19567
QY	1239	ATGGAGGGCAAGGATGAAGAGATCACGAAGACCGAGGCCAGTGCCGGGAGCTGAAGAAG	1298
Db	19566	TTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAG	19507
QY	1299	AAGCTCCAAGAGGAAGAACAACACAGCAAGGAACCTTAGACTAGAAGTGGAGAAGCTGCAG	1358
Db	19506	GAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGGTGGAAGCAAGAGCAGGAG	19447
QY	1359	AAGAGGATGTCTGAGCTGGAGAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGC	1418
Db	19446	GTGGAAGAGCAAGAGCAGGAGCAGGAGCAAGAGCAGGAATTAGAGGAGGTGGAGGAGCAAGAG	19387
QY	1419	ACCAAGTCCATCTGAACCTGGAGAAGGAG	1448
Db	19386	CAGGAGCAGGAGGAGCAGGAGGAGCAGGAG	19357

RESULT 8
US-08-757-669A-20/c
; Sequence 20, Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES' AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-757-669A-20

Query Match	2.1%;	Score 93.6;	DB 3;	Length 32207;
Best Local Similarity	44.6%;	Pred. NO. 1.3e-12;		
Matches 415;	Conservative	0;	Mismatches 509;	Indels 6;
			Gaps	1;
QY	519	GAGCTGGACACACTGGAGGAAAGCAGAAAGGAGACGTACCGCCGCATGCTAGAGCAGCTG	578	

Db 20280 GAGCCACAGCAGCGGGAGCCACAGCAGCGGAGCCACAGCAGCGGGAGCCACAGCAGCGG 20221
QY 579 CTGCTGGCTGAGAAAGTGTACAGGCGCACCGTGTFACGAGCTGGAGAAACGAGAAGCACAAAG 638
Db 20220 GAGCCACAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAG 20161
QY 639 CACACTGACTACATGAACAAGAGCGCAGCAGCTTCACCAACCTGCTGGAGCAGGAGCGAGAG 698
Db 20160 GATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAG 20101
QY 699 AGGTTGAAAAAGCTCTTGAACAAGAAAAAGCTTACCAAGCCCGCAAAAGAAAAAGGAAAAAC 758
Db 20100 GATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAG 20041
QY 759 GCTAAGCGGCTCAACAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCCTTCATG 818
Db 20040 GATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAG 19987
QY 819 TTGGTGGACGAGAGGAGATGCACATCGAGCAACTGGGCCCTGCAGAGTCAAGAAAGTCCAG 878
Db 19986 CAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAG 19927
QY 879 GACCTCACTCAGAAGCTGCTCAAGTTAGAGTGGACTTCAAGCGGTCACTTACAAAATCC 938
Db 19926 CAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAG 19867
QY 939 AAGGAAGACCCAGAGCTGCTCAAGTTAGAGTGGACTTCGAACACAAAGCCTCGAGG 998
Db 19866 GAGCAGGAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAG 19807
QY 999 TTTTCCCAGGACGAGAGATGAACGCCAAATTTGGCGAATCAAGAACTCTCACAAACCGG 1058
Db 19806 CAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAG 19747
QY 1059 CAACCTCGACTCAAACTGGTTGGCTTATCGCAAAGGATTTAGAGGAGCTGGAAGACCAAT 1118
Db 19746 GATGAACAGGAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAG 19687
QY 1119 AAAAGCCTTCAGAAAGCGCAGAGGAGTGAACGCGCAAAATTTGGCGAATCAAGAACTCTCAAAAGGG 1178
Db 19686 GAGCAGGAGCAGGAGTTAGAGGATCAGGAGCAGGAGTTAGAGGAGCAGGAGGAGTTA 19627
QY 1179 GAATGTGGAACCTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTGTCTTGA 1238
Db 19626 GAGGAGCAGGAGCAGGATGAGGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAG 19567
QY 1239 ATGGAGGGCAAGGATGAAGAGATCACGAAGACCGAGGCCCTAGTGCCGGAGCTGAAGAAG 1298
Db 19566 TTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAG 19507
QY 1299 AAGCTCCAAGAGGAAGAACACCAAGCAAGGAACTTAGACTAGAAGTGGAGAAGCTGCGAG 1358
Db 19506 GAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGTTAGAGGAGCAGGAG 19447
QY 1359 AAGAGGATGTCTGAGCTGGAGAAGCTGGAGGAAGCGTTTCAGCCGGAGTTAAGTCGGAATGC 1418
Db 19446 GTGGAAGAGCAAGAGCAGGAGCAGGAAGAGCAGGAATTAGAGGAGGTGGAGGAGCAAGAG 19387
QY 1419 ACCCAGCTCCATCTGAACCTGGAGAAGGAG 1448
Db 19386 CAGGAGCAGGAGGAGCAGGAGGAGCAGGAG 19357

RESULT 9
US-09-230-371A-20/c
; Sequence 20, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A
; APPLICANT: Russo, James J
; APPLICANT: Edelman, Isidore S
; APPLICANT: Moore, Patrick S

; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 45185-G-PCT-US
; CURRENT APPLICATION NUMBER: US/09/230,371A
; CURRENT FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: PCT/US97/13346
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 32207
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-20

Query Match 2.1%; Score 93.6; DB 4; Length 32207;
Best Local Similarity 44.6%; Pred. No. 1.3e-12;
Matches 415; Conservative 0; Mismatches 509; Indels 6; Gaps 1;
QY 519 GAGCTGGACAGACTGGAGGAAAAAGCAGAAAGGAGACGTACCGCCGCATGCTAGAGCAGCTG 578
Db 20280 GAGCCACAGCAGCGGGAGCCACAGCAGCGGGAGCCACAGCAGCGGAGCCACAGCAGCGG 20221
QY 579 CTGCTGGCTGAGAAAGTGTACAGGGCGCACCGTGTACGAGCTGGAGAAACGAGAAGCAACAG 638
Db 20220 GAGCCACAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAG 20161
QY 639 CACACTGACTACATGAACAAGAGCGCAGCAGCTTCACCAACCTGCTGGAGCAGGAGCGAGAG 698
Db 20160 GATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAG 20101
QY 699 AGGTTGAAAAAGCTCCTTGAAACAAGAAAAAGCTTACCAAGCCCGCAAAAGAAAAAGGAAAAAC 758
Db 20100 GATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAG 20041
QY 759 GCTAAGCGGCTCAACAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCCTTCATG 818
Db 20040 GATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAG--GATGAGCAG 19987
QY 819 TTGGTGGACGAGAGGCGCAGATGCACATCGAGCAACTGCGGCTGCAGAGTCAAGAAAGTCCAG 878
Db 19986 CAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAG 19927
QY 879 GACCTCACTCAGAAAGCTGAGGAGGAGGAAGAAAACTCAAAAGCGGTCACTTACAAATCC 938
Db 19926 CAGCAGGATGAGCAGCAGCAGGATGAGCAGGAGCAGCAGGATGAGCAGGAGCAGGATGAG 19867
QY 939 AAGGAAGACCCCGCCAGAAAGCTGCTCAAGTTAGAAGTGAAGTTCGAAACACAAAGGCTCGAGG 998
Db 19866 GAGCAGGAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAG 19807
QY 999 TTTTCCCAGGAGCACGAAGAGATGAACGCGCAAAATTTGGCGAATCAAGAACTCTCAAAATCC 1058
Db 19806 CAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAG 19747
QY 1059 CAACCTTCGACTCAAACTGGTTGGCTTATCGCAAAGGATTTAGAGGAGCTGGAAGACCAAT 1118
Db 19746 GATGAACAGGAGCAGCAGGAGGAGGAGCAGGAGCAGGAGGAGTTAGAGGAGTTAGAG 19687
QY 1119 AAAAGCCTTCAGAAAGCGCAGAGGAGTGAACGCGCAAAATTTGGCGAATCAAGAACTTC 1178
Db 19686 GAGCAGGAGCAGGAGTTAGAGGATCAGGAGCAGGAGTTAGAGGAGCAGGAGGAGTTA 19627
QY 1179 GAATGTGGAACCTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTGTCTTGA 1238
Db 19626 GAGGAGCAGGAGCAGGATGAGGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAG 19567
QY 1239 ATGGAGGGCAAGGATGAAGAGATCACGAAGACCGAGGCCCTAGTGCCGGAGCTGAAGAAG 1298
Db 19566 TTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAG 19507
QY 1299 AAGCTCCAAGAGGAAGAACACCAAGCAAGGAACTTAGACTAGAAGTGGAGAAGCTGCGAG 1358
Db 19506 GAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGTTAGAGGAGCAGGAG 19447
QY 1359 AAGAGGATGTCTGAGCTGGAGAAGCTGGAGGAAGCGTTTCAGCCGGAGTTAAGTCGGAATGC 1418
Db 19446 GTGGAAGAGCAAGAGCAGGAGCAGGAAGAGCAGGAATTAGAGGAGGTGGAGGAGCAAGAG 19387
QY 1419 ACCCAGCTCCATCTGAACCTGGAGAAGGAG 1448
Db 19386 CAGGAGCAGGAGGAGCAGGAGGAGCAGGAG 19357

Db 19506 GAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGGTGGAAGAGCAAGACGAGGAG 19447
QY 1359 AAGAGGATGTCTGAGCTGAGAAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGC 1418
Db 19446 GTGGAAGAGCAAGAGCAGGAGCAGGAAGAGCAGGAATTAGAGGAGGTGGAGGAGCAAGAG 19387
QY 1419 ACCCAGCTCCATCTGAACCTGGAGAAAGGAG 1448
Db 19386 CAGGAGCAGGAGGAGCAGGAGGAGCAGGAG 19357
RESULT 10
US-08-875-435B-2
; Sequence 2, Application US/08875435B
; Patent No. 6593304
; GENERAL INFORMATION:
; APPLICANT: Hasegawa, Kazuhide
; APPLICANT: Arakawa, Emi
; APPLICANT: Oda, Shoji
; APPLICANT: Matsuda, Yuzuru
; APPLICANT: Takahashi, Katsuhito
; APPLICANT: Sugahara, Michihiro
; APPLICANT: Ishiyama, Haruo
; TITLE OF INVENTION: RECOMBINANT DNA COMPRISING DNA CODING
; TITLE OF INVENTION: FOR MYOSIN HEAVY CHAIN SM1 ISOFORM PROTEIN INSERTED INTO
; TITLE OF INVENTION: VECTOR DNA, MICROORGANISM CARRYING THE RECOMBINANT DNA, AND
; TITLE OF INVENTION: AN AGENT FOR TREATMENT OF ARTERIOSCLEROSIS COMPRISING THE
; TITLE OF INVENTION: RECOMBINANT DNA
; FILE REFERENCE: 07898-013001
; CURRENT APPLICATION NUMBER: US/08/875,435B
; CURRENT FILING DATE: 1997-07-25
; PRIOR APPLICATION NUMBER: PCT/JP96/00134
; PRIOR FILING DATE: 1996-01-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 5919
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(5919)
; OTHER INFORMATION: n = A,T,C or G
US-08-875-435B-2
Query Match 1.6%; Score 70.4; DB 4; Length 5919;
Best Local Similarity 37.2%; Pred. No. 3.1e-07;
Matches 413; Conservative 5; Mismatches 680; Indels 12; Gaps 2;
QY 727 AAGCTTACCAAGCCCGCAAGAAAAGGAAACCGCTAAGCGGCTCAACAAACTTCGAGATG 786
Db 4709 AGGCNCTCAANGNCCAGTTNGANVNGATCTCCANGNCGGGANGANCAGAAGAGGAGA 4768
QY 787 AGCTTGTGAAGCTCAAGTCTTCGCCCTCATGTTGGTGGACGAGAGGCAGATGCACATCG 846
Db 4769 AGAGGAGGCANCTNCAGVNCAGCTNCANGAGTANGAGACNGAACTGGAAGANGANCNA 4828
QY 847 AGCAACTGGGCCCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAAGCTGAGGGAGGAG 906
Db 4829 AGCANVNGNCNTGGCNGCAGCAGCNAAGAAAGAGTGGANGGGGACCTTAAAGACCTNG 4888
QY 907 AAGAAAACTCAAAGCGGTCACTTACAAATCCAAGGAAGACCGCCAGAAAGTGTCTCAAGT 966
Db 4889 AGCTNCAGGCNGACTCNGCCATCAANGGGNGGGAGGAGCCATCAAGCAGCTNCNNAAC 4948
QY 967 TAGAAGTGGACTTCGAACACAAGGCGCTCGAGGTTTTCCAGGAGCACGAAGAGATGAACG 1026
Db 4949 TCGAGGCTCAGATGAAGGACTTNCANAGAGANCTGGANGATGCCGTGCCCTCCAGNGANG 5008
QY 1027 CCAAATTGGCGAATCAAGAAATCTCAACACCGGCAACTTCGACTCAAACCTGGTTGGCTTAT 1086
Db 5009 AGATCTTTGCCACNNCNAANGAGAANGAAGAAAGCAAGAGNNTGGANGCAGACCTCA 5068

QY 1087 CGCAAAGGATTGAGGAGCTGGAAGAGAGACCAATAAAGCCCTTCAGAAAGGCAGAGGAAGAGC 1146
Db 5069 TGCAGCTNCAAGAGAGANCTNGCNGCNGCNGAGAGNGCTCGCAANCANGCNGACNTNGAGA 5128
QY 1147 TCCAGGAGCTGAGAGAGAAAATTGCCAAAGGGGAATGTGGAAACTCCAGTCTCATGGCGG 1206
Db 5129 AGGAGGANCTGGCNGAGGAGCTGGCNAGNAGCNTGTCTGGAAGGAANNCTNCAGGANG 5188
QY 1207 AAGTGGAGAGTCT---GCGCAAGCGCGTGTGTTGAGATGGAGGGCAAGGATGAAGAGATCA 1263
Db 5189 AGAAGCGCCGCTGGAGGCNNGGATCGCNCANCTNGAGGAGGAGCTGGAGGANGANCAGG 5248
QY 1264 CGAAGACCGAGGCCAGTGCCGGGAGCTGAAGAAGAAGCTCCAAGAGGAGAAACACCACA 1323
Db 5249 GCAACATGGAGGCNATGAGNGANVNGTNCGCAANGCNCNGCAGGCNGAGCANCNTNA 5308
QY 1324 GCAAGGAACTTAGACTAGAAGTGGAGAAGCTGCAGAAGAGGATGTCTGAGCTGGAGAAGC 1383
Db 5309 GCAANGAGCTGGCCACAGANCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 5368
QY 1384 TGGAGGAAGCTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGA 1443
Db 5369 TNGAG-----CGNCAGAAACAAGGANCTNMRNAGCAAGNTNCANGANTNGANGGNG 5419
QY 1444 AGGAGAAGAACCTAACCAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGAGTTA 1503
Db 5420 CNGTCAANNCCAGNTCAANTCCACNNTNGCGCGCTGGAGGCCAAGATTGCNCAGCTNG 5479
QY 1504 AAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAGCGCGAGTTAAGCCTCAAAGATGACC 1563
Db 5480 AGGAGCAGGTNGANCAGGAGGCCAGAGAGAAAGCAGGCGGCCNCCAGNCGCTGAAGCANA 5539
QY 1564 TTACAAAGCTGAGTCTTCACTGTGATGCTGGTGGATGAGAGGAGAAATATATATGAGGA 1623
Db 5540 NNGACAAGAAAGCTNAAGGANNTNTGCTGCAGGTGGANGANGAGCGCAAGATGCCNGAGC 5599
QY 1624 AAATAAGCAAGAGAGAGGAAAGTGGATGGGTTGAATAAAAACTTTAAGGTGGAGCAGG 1683
Db 5600 AGTACAAGGAGCAGGAGGAGGAGGAAAGNANNCCANGGTCAAGCAGCTNAAGAGCAGCTGG 5659
QY 1684 GAAAAGTCATGGATGTGACGGAAAAAGCTAATCGAGGAGAAAGCAAGAAAGCTTTTAAAACTCA 1743
Db 5660 ANGAGGCGNGAGGAGGAGTNCAGNGCATCAACGCCCAACCGCAGGAAGCTGCAGCGGGAGC 5719
QY 1744 AATCTGAAATGGAGGAGGAGGAGTACAGTCTGCACAAAGGAGGAGGATGAGCTGATGGGTA 1803
Db 5720 TNGANGAGGCCACNGAGAGCAANGAGGCGCCATGGGCCGNGAGGTGAACGCNCTCAAGAGCA 5779
QY 1804 AACTGAGGAGCGAAGAAAGGTCCTGTG 1833
Db 5780 ANCTCAGGVGAGGAAACGAGNCNTCNTNG 5809
RESULT 11
US-09-249-585A-2
; Sequence 2, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISODES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1926)
; OTHER INFORMATION: coding strand of EBNA-1 DNA

US-09 -249 -585A-2					
Query Match		1.5%;	Score 65.6;	DB 4;	Length 1926;
Best Local Similarity		44.2%;	Pred. No. 2.7e-06;		
Matches 269; Conservative		0;	Mismatches 339;	Indels	0; Gaps 0;
QY	846	GAGCAACTGGCCTGCAGAGTCAGAAAGTCCAGGACCTCTACTCAGAAGCTGAGGGAGGAG	905		
Dd	302	GAGGGGCAGGACGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG	361		
QY	906	GAAGAAAAACTCAAAGCGGTCACTTACAATCCAAGGAAGACGCCAGAAAGCTGCTCAAG	965		
Dd	362	CAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAG	421		
QY	966	TTAGAACTGGACTTCGAACACAAGGCCTCGAGGTTTTCCACAGGAGCACGAAGAGATGAAC	1025		
Dd	422	GGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG	481		
QY	1026	GCCAAATTGGCGAATCAAGAAATCTCAACCGGCAACTTCGACTCAAACTGGTTGGCTTA	1085		
Dd	482	GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGAGGGGCAGGAGGGG	541		
QY	1086	TCGCAAGGATTGAGGAGCTGGAAGAGACCATAAAGCCCTTCAGAAGGCAGAGGAAGAG	1145		
Dd	542	CAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGAGGAG	601		
QY	1146	CTCCAGGAGCTGAGAGAGAAAATTGCCAAGGGGAATGTGGAAACTCCAGTCTCATGGCG	1205		
Dd	602	GGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAG	661		
QY	1206	GAA GTGGAGTCTGCGCAAGCGCGTGCTTTCAGATGGAGGCAAGGATGAAGAGATCACG	1265		
Dd	662	GAGGAGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGAGGGGCAG	721		
QY	1266	AAGACCGAGGCCAGTGCCGGGAGCTGAAGAAGAAGCTCCAAGAGGAAGAACACCACAGC	1325		
Dd	722	GAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGG	781		
QY	1326	AAGGAAC TTAGACTAGAA GTGGAGAAAGCTGCAGAAAGAGGATGTCTGAGCTGGAGAAGCTG	1385		
Dd	782	CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGAGGGGCAGGAGGGGCAGGAGCAG	841		
QY	1386	GAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAG	1445		
Dd	842	GAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAG	901		
QY	1446	GAGAAGAA	1453		
Dd	902	GAGGGGCA	909		

```

RESULT 12
US-09-410-399-3
; Sequence 3, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; APPLICANT: Cotter, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; TITLE OF INVENTION: to Genomic Host DNA
; FILE REFERENCE: UM-03778
; CURRENT APPLICATION NUMBER: US/09/410,399
; CURRENT FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein-Barr virus
US-09-410-399-3

```

RESULT 13
 US-09-050-863-2
 ; Sequence 2, Application US/09050863
 ; Patent No. 611411
 ; GENERAL INFORMATION:
 ; APPLICANT: Lao, Ying
 ; APPLICANT: Hiang, Betty
 ; APPLICANT: Payan, Don
 ; TITLE OF INVENTION: Mammalian Protein Interaction Cloning
 ; TITLE OF INVENTION: System
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/050,863
 ; FILING DATE: 30-MAR-1998

```

; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-09-050-863-2

Query Match      1.5%; Score 65.6; DB 3; Length 2580;
Best Local Similarity 44.2%; Pred. No. 3.1e-06;
Matches 269; Conservative 0; Mismatches 339; Indels 0; Gaps 0;

QY      846 GAGCAACTGGCCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAAGCTGAGGGAGGAG 905
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      685 GAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGG 744

QY      906 GAAGAAAACTCAAAGCGTCACTTACAAATCCAAGGAAGACCGCCAGAACTGCTCAAG 965
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      745 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAG 804

QY      966 TTAGAAGTGGACTTCGAACACAAAGGCCTCGAGGTTTTCCAGGAGCACGAAGAGATGAAC 1025
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      805 GGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGAGGAG 864

QY      1026 GCCAAATTGGCGAATCAAGAATCTCAACCCGGCACTTCGACTCAAACCTGTTGGCTTA 1085
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      865 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGG 924

QY      1086 TCGCAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAAGGCAGAGAAAGAG 1145
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      925 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAG 984

QY      1146 CTCCAGGAGCTGAGAGAGAAAATTGCCAAAGGGGAATGTGGAAACTCCAGTCTCATGGCG 1205
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      985 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 1044

QY      1206 GAAGTGGAGAGTCTCGCAAGCGCGTCTCACAACCGGCAACTTCGACTCAAACCTGTTGGCTTA 1265
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1045 GAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 1104

QY      1266 AAGACCGAGGCCAGTGCCGGGAGCTGAAGAAGAAAGTCCAAGAGGAAGAACACACAGC 1325
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1105 GAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAG 1164

QY      1326 AAGGAACTTAGACTAGAACTGGAAGAGCTGCAGAAAGAGGATGTCTGAGCTGGAGAAGCTG 1385
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1165 CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 1224

QY      1386 GAGGAAGCGTTTCAGCCGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAG 1445
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1225 GAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 1284

QY      1446 GAGAAGAA 1453
Db      ||| ||| |||
Db      1285 GAGGGGCA 1292
```

RESULT 14
US-09-359-081-2
; Sequence 2, Application US/09359081
; Patent No. 6316223
; GENERAL INFORMATION:
; APPLICANT: Lao, Ying
; Hiang, Betty

```

; Payan, Don
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning
; System
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/359,081
; FILING DATE: 22-Jul-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/050,863
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-359-081-2
```

```

Query Match      1.5%; Score 65.6; DB 4; Length 2580;
Best Local Similarity 44.2%; Pred. No. 3.1e-06;
Matches 269; Conservative 0; Mismatches 339; Indels 0; Gaps 0;

QY      846 GAGCAACTGGCCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAAGCTGAGGGAGGAG 905
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      685 GAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGG 744

QY      906 GAAGAAAACTCAAAGCGTCACTTACAAATCCAAGGAAGACCGCCAGAAAGCTGCTCAAG 965
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      745 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAG 804

QY      966 TTAGAAGTGGACTTCGAACACAAAGGCCTCGAGGTTTTCCAGGAGCACGAAGAGATGAAC 1025
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      805 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAG 864

QY      1026 GCCAAATTGGCGAATCAAGAATCTCACAACCGGCAACTTCGACTCAAACCTGTTGGCTTA 1085
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      865 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGG 924

QY      1086 TCGCAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAAGGCAGAGAAAGAG 1145
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      925 CAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGAG 984

QY      1146 CTCCAGGAGCTGAGAGAGAAAATTGCCAAAGGGGAATGTGGAAACTCCAGTCTCATGGCG 1205
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      985 GGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAG 1044

QY      1206 GAAGTGGAGAGTCTCGCAAGCGCGTCTCACAACCGGCAAGGATGAAGAGATCAAG 1265
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1045 GAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 1104

QY      1266 AAGACCGAGGCCAGTGCCGGGAGCTGAAGAAGAAAGTCCAAGAGGAAGAACACACAGC 1325
```

Db	1105	GAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGG	1164
QY	1326	AAGGAACCTTAGACTAGAACTGGAGAACTGCAGAAAGGATGTCTGAGCTGGAGAAGCTG	1385
Db	1165	CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAG	1224
QY	1386	GAGGAAGCGTTCAGCCGGACTAAGTCGGGAATGCACCCAGCTCCATCTGNAACCTGGAGAAG	1445
Db	1225	GAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGAGGGGCAG	1284
QY	1446	GAGAAGAA	1453
Db	1285	GAGGGGCA	1292

```

RESULT 15
US-09-130-114-1/c
; Sequence 1, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/1D903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5452
; TYPE: DNA
; ORGANISM: VEBNA
US-09-130-114-1

```

Query Match	1.5%;	Score	65.6;	DB	2;	Length	5452;
Best Local Similarity	44.2%;	Pred. No.	4.6e-06;				
Matches	269;	Conservative	0;	Mismatches	339;	Indels	0;
Gaps	0;						
QY	846	GAGCAACTGGGCCCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAGCTGAGGGAGGAG	905				
Db	2120	GAGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG	2061				
QY	906	GAAGAAAACTCAAAGCGGTCACTTACAATCCAAGGAAGACCGCCAGAAAGCTGCTCAAG	965				
Db	2060	CAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAG	2001				
QY	966	TTAGAAGTGGACTTCGAACACAAGGCCTCGAGGTTTTTCCAGGAGCACGAAGAGATGAAC	1025				
Db	2000	GGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAG	1941				
QY	1026	GCCAAATTGGCGAATCAAGAATCTCACAAACCGGCAACTTCGACTCAAACCTGGTTGGCTTA	1085				
Db	1940	GAGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGAGGGGCAGGAG	1881				
QY	1086	TCGCAAAGGATTGAGAGCTGGAAGAGACCAATAAAGCCTTCAGAAAGCAGAGGAAGAG	1145				
Db	1880	CAGAGGGGCAGGACAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAG	1821				
QY	1146	CTCAGGAGCTGAGAGAGAAAATTGCCAAAGGGGAATGTGAAACTCCAGTCTCATGGCG	1205				
Db	1820	GGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAG	1761				
QY	1206	GAAGTGAGAGTCTGCGCAAGCGCTGCTTGAGATGAGAGGGCAAGGATGAAGAGATCACG	1265				
Db	1760	GAGGAGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAG	1701				
QY	1266	AACACGAGGCCCACTGCCGGGAGCTGAAGAAGAGCTCCAAGAGGAAGAACCCACAGC	1325				
Db	1700	GACAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGG	1641				

Qy	1326	AAGGAACCTAGACTAGAAAGTGGAGAAAGCTGCAGAAAGAGGATGTCTGAGCTGGAGAAGCTG	1385
Db	1326		
	1640	CAGGAGCAGGAGGGGCAGAGGGGCAGGAGCAGGAGGGGCAGGAGGGCAGGAGCAG	1581
Qy	1386	GAGGAACGCTTCAGCCGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAG	1445
	1386		
Db	1580	GAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAG	1521
	1580		
Qy	1446	GAGAAGAA	1453
Db	1446		
	1520	GAGGGGCA	1513

Search completed: September 8, 2004, 11:04:06
Job time : 329 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: September 7, 2004, 13:26:35 ; Search time 137.5 Seconds
(without alignments)
6105.888 Million cell updates/sec

Title: US-10-788-793-1
Perfect score: 7731
Sequence: 1 ccactgggtttcttcaaggga.....aaaaaaaaaaaaaaaaaaaa 4364

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10788793/runat_07092004_134022_10821/app_query.fasta_1.4551
-DB=PIR_78 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=Bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10788793 @CGN 1 1 218 @runat_07092004_134022_10821 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				%	
Result No.	Score	Query Match	Length DB ID	Description	
1	479	6.2	2139 2 T18296	myosin heavy chain	
2	476	6.2	2094 2 S33124	tpr protein - huma	
3	473.5	6.1	2663 1 S28261	centromere protein	
4	471.5	6.1	2007 1 B43402	myosin heavy chain	
5	469.5	6.1	1961 1 A61231	myosin heavy chain	
6	467	6.0	1909 2 A45592	liver stage antige	
7	464.5	6.0	1790 2 S67593	transport protein	
8	463.5	6.0	2020 2 T21174	hypothetical prote	
9	462.5	6.0	1164 2 T24806	hypothetical prote	
10	461.5	6.0	1999 1 S21801	myosin heavy chain	
11	457	5.9	1938 1 A40997	myosin heavy chain	
12	453.5	5.9	1992 2 A47297	myosin heavy chain	
13	450.5	5.8	1392 2 A43336	microtubule-vesicl	
14	449	5.8	1938 2 JC5421	smooth muscle myos	

15	449	5.8	1972	2	JC5420	smooth muscle myos
16	448.5	5.8	1964	2	A59282	nonmuscle myosin I
17	447.5	5.8	1959	1	A33977	myosin heavy chain
18	447	5.8	1427	2	S22695	restin - human
19	446.5	5.8	1039	2	S18199	myosin heavy chain
20	445.5	5.8	1407	1	S28589	trichohyalin - rab
21	445	5.8	1972	1	A41604	myosin heavy chain
22	443	5.7	2017	1	A36014	myosin heavy chain
23	443	5.7	2057	2	S61477	myosin II heavy ch
24	441.5	5.7	2253	2	T30336	nuclear/mitotic ap
25	441	5.7	880	2	F75103	conserved hypothet
26	441	5.7	1979	1	S03166	myosin heavy chain
27	439	5.7	1690	2	T13030	microtubule bindin
28	439	5.7	1827	2	T16270	hypothetical prote
29	439	5.7	1940	1	A24922	myosin heavy chain
30	438	5.7	1976	2	A59252	myosin heavy chain
31	438	5.7	2346	2	T13829	Tpr homolog - frui
32	435	5.6	1133	2	T22976	hypothetical prote
33	434	5.6	1410	1	A57013	early endosome ant
34	433.5	5.6	955	2	S24348	myosin heavy chain
35	433.5	5.6	1597	2	S68420	citron - mouse
36	433.5	5.6	1927	2	A59236	embryonic muscle m
37	433	5.6	1938	1	JX0178	myosin heavy chain
38	431	5.6	2116	2	A26655	myosin heavy chain
39	430.5	5.6	1937	2	I38055	myosin heavy chain
40	430	5.6	1931	2	A59234	slow myosin heavy
41	429	5.5	1898	1	A45973	trichohyalin - hum
42	429	5.5	1940	2	A59287	myosin heavy chain
43	429	5.5	3259	1	A56539	giantin - human
44	428.5	5.5	1940	1	S04090	myosin heavy chain
45	428	5.5	4687	1	A39638	plectin - rat

ALIGNMENTS

RESULT 1
T18296
myosin heavy chain - Entamoeba histolytica
C;Species: Entamoeba histolytica
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C;Accession: T18296
R;Guillen, N.
submitted to the EMBL Data Library, February 1997
A;Reference number: Z18865
A;Accession: T18296
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2139 <GUI>
A;Cross-references: EMBL:L03534; NID:g1850912; PID:g1850913; PIDN:AAB48065.1
C;Genetics:
A;Gene: mhca
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;91-780/Domain: myosin motor domain homology <NMO>

Alignment Scores:
Pred. No.: 6.47e-15 Length: 2139
Score: 479.00 Matches: 228
Percent Similarity: 42.55% Conservative: 189
Best Local Similarity: 23.27% Mismatches: 350
Query Match: 6.20% Indels: 214
DB: 2 Gaps: 36

US-10-788-793-1 (1-4364) x T18296 (1-2139)

QY	174	GAAGATGCAAAAAAGAAC-----AAGGCCAATCGGAAGGAGGAGGATGTC	218
Db	871	GluspleuLysLysLysLeuAlaGluGluLysLysArgGluAlaGluAla	890
QY	219	ATGGCTTCGGGAAGTATCAAAAGGCACCTCAACCATCTGGAGAAAGACT	278
		:::	:::
Db	891	LeuAlaSerAlaThrAlaLys-----ThrGlyGluLeuGluAlaLysIle	905
QY	279	AAGAAGTCTGTGGAGTTATCCAAGGAGGACCTCATCCAGCTCTCTGAGTATCATGGAAGGG	338

Db 906 -----GlnAspLeuGluAspLysIleSerGluLeuGluSer 917
QY 339 GAGTTGCAGGCTCGAAGAAGATGTATCCACATGCTGAGGACAGAGAAAAACCAAGCCCGAG 398
Db 918 LysLeuSerAlaAlaGlu-----LeuAspLysGlnGluLeuAsnLeuLys 932
QY 399 GTTCTGGAGGCACACTATGGATCTGCAGAACCTGAGAAAAGTGCTTCGGGTCTCGCACCGA 458
Db 933 ileGluAsnLeuGluGluAspLysGluGluLeuLysGluThrIleAspAsnLeuLysGly 952
QY 459 GATGCCATCCTTGCTCAAGAGAAAGTCCATAGGAGAAGACGCTCTATGAGAAAACCTATCTCA 518
Db 953 Asp-----LeuLysAspSerLysLeuLysGlyGluAspLeu--GluValGluIleThr 969
QY 519 GAGCTGGACAGACTGGAGGAAAAGCAGAGAGACGTTACCGCCGCATGCTAGAGCAGCTG 578
Db 970 GluLeuAsn-----SerGlnIleAsnThrLeuAsnAlaThrValAspLys--- 984
QY 579 CTGCTGGCTGAGAAGTGTCTACAGCGCCACCGTGTACGAGCTGGAGAACCGAGAAGCACAAAG 638
Db 985 -----LysAspLysThrIleAlaGluMetGlnGluSerIleAspGlu 998
QY 639 CACACTGACTACATGAACAAGAGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGAGAG 698
Db 999 LysGluAspGluIleThrLysLeuLysGlyAspIleLysLeuLeuGluGluLysAsp 1018
QY 699 AGGTTGAAAAGCTCCTTGACAAAGAAAAGCTTACCAGCCCGCAAAGAAAAGGAAAAC 758
Db 1019 Asp-----LeuGluGlnAspArgAlaAspValSerAlaThrLysAspAspIle 1034
QY 759 GCTAAGCGGCTCAACAAACTT-----CGAGATGAGCTTGTGAAG 797
Db 1035 AlaLysLeuAsnLysIleThrIleGluCysGluAspAlaLysAspGluIleAlaLys 1054
QY 798 CTCAAGTCTTCGCCCTCATGTTGGTGGACGAGGCAGATGCACATCGACCAACTGGGC 857
Db 1055 LeuGlu-----GlnGluLeuGlu 1060
QY 858 CTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAAGCTGAGGGAGGAGAAACATC 917
Db 1061 AspGluGluAsnLysAsnLysAspLeuThrAsnGluLeuGlnGlnThrGlnLeuLysLeu 1080
QY 918 -----AAAGCGGTCACTTACAAATCCAAAGGAAGAC 947
Db 1081 GlyGluThrGluLysSerLeuAlaAlaGlnValAlaAlaThrLysLysAlaSerAspGlu 1100
QY 948 CGCCAGAAAGCTGCTCAAGTTAGAACTGGACTTCGAACACAAGGCCTCGAGGTTTTCCACAG 1007
Db 1101 ArgAspThrLeuSerGln-----AsnLeu 1108
QY 1008 GAGCACGAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCACAAACCGGCAACTTCGA 1067
Db 1109 GluAsnGluLysLeuThrThrLys-----AsnLeuThrLysThrLysAlaAspLeuGlu 1126
QY 1068 CTCAAACTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGAA---GAGACCAATAAA--- 1121
Db 1127 LysLysIleSerGlyLeuLysGlnAspTyrGluAspLeuGluAspAspLysAsnLysIle 1146
QY 1122 -----AGCCTTCAGAAAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAATTGCCCAA 1175
Db 1147 GluGlyAspLeuArgAsnAlaGlnArgLysIleLysGluLeuAspAspGluIleThrLys 1166
QY 1176 GGGGAATGTGGAAACTCCAGTCTCATGCGGGAAGTGGAGAGTCTGCGCAAGCGCTGCTT 1235
Db 1167 GlyAlaAspValSerGlnTyrLeuGlnLysGlnLysGluGluTyrGluSerGlnIleAla 1186
QY 1236 GAGATGGAG-----GGCAAGGATGAAGAGATC 1262
Db 1187 LysMetGlnGluGluLysGluAlaIleGlyAsnAspValLysAsnLysGluLysThrIle 1206
QY 1263 ACGAAGACCGAGGCCCTGCGGGAGCTGAAGAAGAGCTCCAAGAG-----GAA 1313

Db 1207 LysGluLysGluLeuGluIleGlnSerLeuGlnGluLysLeuAspGluThrGluValGlu 1226
QY 1314 GAACACACAGCAAGGAACCTTAGACTAGAAAGTGGAGAAAGCTGCAGAAAGGATG---TCT 1370
Db 1227 LysGluAspAlaGluLysLysLysGluIleGluLysGluMetLysAlaLeuGlnGlu 1246
QY 1371 GAGCTGGAGAAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCAT 1430
Db 1247 GluLysGluAsnValGluSerSerLysAsnSerThrGluLysAspLysLysLeuGlu 1266
QY 1431 CTGAACCTGGAGAAAGGAGAAACCTAACCAAGACCTGCTGAACGAGCTGGAGGTGGTC 1490
Db 1267 AspAsnLeuLysAspThrGlnLysLysLeuAspAspMetThrAlaAspAsnGluLysLeu 1286
QY 1491 AAGAGTCGAGTTAAAGAACTCGAATGCTCC-----GAGAGT 1526
Db 1287 LysAlaLysAlaLysAspLeuGluAlaGlnLeuAsnGluValGlnAspAsnHisGluLys 1306
QY 1527 AGACTGGAGAAAGCGCGAGTTA-----ATGATGGAGAAAAATAAAG 1547
Db 1307 AlaValAlaAspAlaGluLeuLeuAsnLysLysLysAlaGlnSerAspLysGluLeuAsn 1326
QY 1548 AGCCTCAAA-----GATGACCTTACAAAAGCTGAAGTCCCTTCACCTGTGATGCTGGTG 1598
Db 1327 SerLeuLysAlaGluLeuGluAlaLeuThrLysAlaLysSer-----Val 1341
QY 1599 GATGAGAGGAAAAAT-----ATGATGGAGAAAAATAAAG 1631
Db 1342 ValGluSerLysAsnLysAspSerGluAsnGluLysAlaAlaLeuSerGluGluIleAsp 1361
QY 1632 CAAGAAGAGAGGAAAGTGGATGGTTGAATAAAAACTTTAAAGGTGGAGCGGAAAAAGTC 1691
Db 1362 GlnAlaAsnGluLysLeuLysAsnIleGlnAlaAspLeuArgLysAlaThrAlaAspLeu 1381
QY 1692 ATGGATGTGACGGAA-----AAGCTAATCGAGGAA 1721
Db 1382 GlnGluAlaAsnGluLysLysAlaGluValGluAlaGlnArgAspLysLeuValAlaAsp 1401
QY 1722 AGCAAGAAGCTTTTAAAACTCAAATCTGAAATCGAGGAAAAAGGAGTACAGTCTGACAAAG 1781
Db 1402 AsnLysLysMetThrLysThrLeuGluGluIleLysAlaArgAspGluAsnThrTyr 1421
QY 1782 GAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGGAAGAAAGAAAGTCTGTGAAGTGAAGC 1841
Db 1422 LysValGluAsnTyrGluLysValLeuLysArgLysGluAlaAspLeuGluAlaAsn 1441
QY 1842 TGCAGTGTAGACTTACTAAAG-----AAGCGGCTT 1871
Db 1442 GluAsnLeuAspIleGluLysLysAspArgMetAsnLysGluLysGlnValLysLysLeu 1461
QY 1872 GATGGC---ATAGAGGAGGTAGAAAGGAAATAAACCCAGGTAGTCTGTCGAAGGGGTCT 1928
Db 1462 GluGlyGluLeuLysGluThrLysAspLysLeuAsnAlaIleAlaGluLysAspSer 1481
QY 1929 GAGTTCACCTGCCCCGAA-----GACAATAAGATCAGAGAACTAACGCTTGAAATCGAG 1982
Db 1482 IlePheThrAlaLysLysGlnSerAspAlaAspLeuGluGluLeuAsnLysThrValGlu 1501
QY 1983 AGA-----CTGAAGAAACCGGCTCCAGCAGTTGGAGGTGGTGGAGGGGACTTG 2030
Db 1502 GluHisAspGluValValAlaLysLeuAsnThrGlnIleThrLysLeuThrArgAspAsn 1521
QY 2031 ATGAACACCGAGGACGAATATGACCAGTTGGAGCAGAGATTTCAGAACCGAGCAGGATAAG 2090
Db 1522 GlnSerAlaGluGluLeuAsnGluLeuArgSerLysAlaAspLysAspLysLysLys 1541
QY 2091 GCAAACTCTCTCCAGCAGCTCGAGGAAATCAAAACACCAAAATGGCCAAAGCACAAAGCC 2150
Db 1542 IleSerGluLeuGluGlnValAsnGluLeuGluSerArgProValGlyThrGlyAsn 1561
QY 2151 ATAGAGAAAGGGGAGGCCGTG-----AGCCAGGAAGCCGAACCTGGACACAGGTTT 2201
Db 1562 AlaAspGluAsnGluIleLysIleArgAspAlaGlnIleAlaAspLeuAsnLysAlaLeu 1581

QY 2202 CGGCTGGAGGAGGCTAAAGTCGTGATTATACAGGCCGAGGTGCAGGCTCTCAAGGAGAAG 2261
Db 1582 GluMetLysGlyValGlnAsnAsnGlnLeuGlnAlaThrAsnLysGluLeuLysAlaLys 1601
QY 2262 ATCCACGAGCTGATGAACAAG----- 2288
Db 1602 AspAsnAspLeuThrSerLysIleGluIleThrGluAsnGluMetLysLysLeuGluAsn 1621
QY 2289 CAGCTGTCTCAGCTCCAAGTCGACTATTCCGGTCCTTCAGCAAAAGATTTATGGAAGAAGAA 2348
Db 1622 AlaLysLysArgLeuGluGlnAspLysAspGluAlaAspLysAlaValSerGluGlnThr 1641
QY 2349 ACTAAGAACAAGACATGGGAGGGAGGTCTCAATCTGACCAAGGAGCTAGAGCT---- 2404
Db 1642 IleLysArgLysGlyLeuGluGluValLysLysLeuThrThrGluIleGlnAlaLeu 1661
QY 2405 ---TTCCAAGCGCTACAGCGGAGCTCTCAGGCCGAGTGGGAACGCCGAGGATGGTGA 2461
Db 1662 LysPheGlnIleAsnAlaProSerSerValAlaGlnGluGlu-LysGlnArgLeuGl 1681
QY 2462 CGTGCCTGTGGCCTCCACTCGGGTGCAGACCCGAGCGGTGTGCGGGATGCTGCGGAGGA 2521
Db 1681 uSerAspIleAlaGluLeuLysGluGlnLeuGluGlnGluArgThrThrAlaAlaAsnAl 1701
QY 2522 GGAGACCCCGCTGTTCATTCCGAAATCCTTCCAGGAGAA---AATCACATCATGAG 2578
Db 1701 aGluAlaGlu-----ArgLysLysIleGlnAlaGluLeuAspGluValLysPh 1717
QY 2579 TAATCTTCGACAGGTAGGC-----CTGAAGAAACCCCATGGAACGGTC 2620
Db 1717 eAsnLeuGluAspValThrAsnGlnArgGluLysLeuValAlaLysAsnSerGluAsnAs 1737
QY 2621 CTCGGTCTCGACAGGTATCCCCCAGCAGCGAATGAGTCAACATGAGGAAGTC----- 2674
Db 1737 pAlaGluIleAsp-SerLeuLysGluGluLysLys-AlaLeuGluAspGluIleGluLys 1756
QY 2675 -----TTGGATTCTTGGATGAGAAA 2695
Db 1757 IleThrAspAspAsnAsnLysLeuSerGluGluIleAspSerLeuAspArgLys 1774

RESULT 2
S33124
tpr protein - human
N;Alternate names: kinase-related transforming protein (tpr-met); protein with promoter
C;Species: Homo sapiens (man)
C;Date: 22-Nov-1993 #sequence revision 26-May-1995 #text_change 11-Jan-2002
C;Accession: S33124; S23740; S00928; G01185
R;Mitchell, P.J.; Cooper, C.S.
Oncogene 7, 2329-2333, 1992
A;Title: The human tpr gene encodes a protein of 2094 amino acids that has extensive coi
A;Reference number: S33124; MUID:93064711; PMID:1437155
A;Accession: S33124
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-2094 <MIT>
A;Cross-references: EMBL:X66397; NID:g633225
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991
R;Mitchell, P.J.; Cooper, C.S.
Oncogene 7, 383-388, 1992
A;Title: Nucleotide sequence analysis of human tpr cDNA clones.
A;Reference number: S23740; MUID:92195670; PMID:1549355
A;Accession: S23740
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-725,'L' <MI2>
A;Cross-references: EMBL:X63105; NID:g37257; PIDN:CAA44819.1; PID:g37258
R;King, H.W.S.; Tempest, P.R.; Merrifield, K.R.; Rance, A.J.
Oncogene 2, 617-619, 1988
A;Title: Tpr homologues activate met and raf.
A;Reference number: S00928; MUID:88262257; PMID:3387099
A;Accession: S00928
A;Molecule type: mRNA

A;Residues: 1-31,'R',33-142 <KIN>
A;Cross-references: EMBL:Y00672; NID:g37255; PIDN:CAA68681.1; PID:g37256
R;Greco, A.
submitted to the EMBL Data Library, December 1995
A;Reference number: H00592
A;Accession: G01185
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 144-228 <GRE>
A;Cross-references: EMBL:X94208; NID:g1296797; PIDN:CAA63904.1; PID:g1296798
C;Genetics:
A;Gene: GDB:TPR
A;Cross-references: GDB:128821; OMIM:189940
A;Map position: 1q25-1q25
A;Introns: 177/3

Alignment Scores:
pred. No.: 8.88e-15 Length: 2094
Score: 476.00 Matches: 314
Percent Similarity: 36.85% Conservative: 252
Best Local Similarity: 20.44% Mismatches: 510
Query Match: 6.16% Indels: 460
DB: 2 Gaps: 67

US-10-788-793-1 (1-4364) x S33124 (1-2094)

QY 174 GAAGATGCAAAAAGAACAAAGCCCAATCGGAAGGAGGAG-----GATGTCATG 221
Db 508 GluGluAlaArgGlyAsnHisValIleArgAspGluValSerSerAlaAspIleSer 527
QY 222 GCTTCCGGA---ACTATCAAAAGGCAC----- 245
Db 528 SerSerSerGluValIleSerGlnHisLeuValSerTyrArgAsnIleGluGluLeuGln 547
QY 246 -----CTCAAAACCATCTGGAGAAAGTGAGAAAAGACT 278
Db 548 GlnGlnAsnGlnArgLeuLeuValAlaLeuArgGluLeuGlyGluThrArgGluArgGlu 567
QY 279 AAGAAGTCTGTGGAGTTATCCAAG----- 302
Db 568 GluGlnGluThrThrSerSerLysIleThrGluLeuGlnLeuLysLeuGluSerAlaLeu 587
QY 303 GAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGGAGTTGCAGGCTCGAGAAAGATGTC 362
Db 588 ThrGluLeuGluGlnLeuArgLysSerArgGlnHisGlnMetGlnLeuValAspSerIle 607
QY 363 ATC-----CACATGCTGAGGACAGAGAAACCAAGCCCGAGGTCTGGAG 407
Db 608 ValArgGlnArgAspMetTyrArgIleLeuLeuSerGlnThrThrGlyValAlaIlePro 627
QY 408 GCACACTATGGATCTGCAGAA----- 428
Db 628 LeuHisAlaSerSerLeuAspAspValSerLeuAlaSerThrProLysArgProSerThr 647
QY 429 -----CCTGAGAAAGTGCTT-----CGGGTCTGCAC 455
Db 648 SerGlnThrValSerThrProAlaProValProValIleGluSerThrGluAlaIleGlu 667
QY 456 CGAGATGCCATCCTTGCTCAAGAGAAGTCCATAGGAGAAGACGTCCTATGAGAAACCTATC 515
Db 668 AlaLysAlaAlaLeuLysGlnLeuGlnGluIlePheGluAsn--TyrLysLysGluLys 686
QY 516 TCAGAGCTGGACAGACTGGAGGAAAGCAGAGGAGACGTACCGC----- 560
Db 687 AlaGluAsnGluLysIleGlnAsnGluGlnLeuGluLysLeuGlnGlnValThrAsp 706
QY 561 -----CGCATGCTAGACGACCTGCTGCTGGCTGAGAAG----- 593
Db 707 LeuArgSerGlnAsnThrLysIleSerThrGlnLeuAspPheAlaSerLysArgTyrGlu 726
QY 594 -----TGTCACAGGCGCACCGTGTACGAGCTGGAGACGAGAAG 632
Db 727 MetLeuGlnAspAsnValGluGlyTyrArgArgGluIleThrSerLeuHisGluArgAsn 746

QY	633	CACAAAGCACACTGACTACATGAACAACAGAGCGACGACACTTCAACCAACCTGCTGGAGCAGGAG	692
Db	747	GlnLysLeuThrAlaThrThrGlnLysGlnGluGlnIleIleAsnThrMetThrGlnAsp	766
QY	693	--CGAGAGAGGTTGAATAAAGCTCCTTGAACAAGAA---AAAGCTTACCAAGCCCGCAA	746
Db	767	LeuArgGlyAlaAsnGluLysLeuAlaValAlaGluValArgAlaGluAsnLeuLysLys	786
QY	747	GAATAAGGAAAAACGCTAAG-----CGGCTCAACAAAACCTTCGAGATGAGCTTGTG	794
Db	787	GluLysGluMetLeuLysLeuSerGluValArgLeuSerGlnGlnArgGluSerLeu---	805
QY	795	AAGCTCAAGTCCTTCGCCCTCATGTTGGTGGACGAGAGGCGAGATGCACATCGAGCAACTG	854
Db	806	-----LeuAlaGluGlnArgGlyGlnAsnLeuLeuThr	817
QY	855	GGCCTGCAGAGT-----CAGAAAAGTC	875
Db	818	AsnLeuGlnThrIleGlnGlyIleLeuGluArgSerGluThrGluThrLysGlnArgLeu	837
QY	876	CAGGACCTCACTCAGAAGCTGAGGAGGAGGAGAAAGAAAACCTCAAAGCGCTCACTTACAA	935
Db	838	SerSerGlnIleGluLysLeuGluHisGluIleSerHisLeuLysLysLysLeuGluAsn	857
QY	936	TCCAAGGAACCGCCAGAAAGCTGCTCAAG---TTAGAAGTGGAC---TTCGAACACACAAG	989
Db	858	GluValGluGlnArgHisThrLeuThrArgAsnLeuAspValGlnLeuLeuAspThrLys	877
QY	990	GCCTCGAGGTTTCCCAGGAGCAGCAAGAGATGAACGCCAAA-----	1031
Db	878	ArgGlnLeuAspThrGluThrAsnLeuHisLeuAsnThrLysGluLeuLysAsnAla	897
QY	1032	-----TTGGCGAATCAAGAAATCTCACAAACCGGCAACTTCGACTCAAACCTGGTTGGC	1082
Db	898	GlnLysGluIleAlaThrLeuLysGlnHisLeuSerAsnMetGluValGlnValAlaSer	917
QY	1083	---TTATCGCAAAGGATT-----GAGGAGCTGGAAGAGACC	1115
Db	918	GlnSerSerGlnArgThrGlyLysGlyGlnProSerAsnLysGluAspValAspAspLeu	937
QY	1116	AATAAAGCCTTCAGAAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAATTGCCAAA	1175
Db	938	ValSerGlnLeuArgGlnThrGluGluGlnValAsnAspLeuLysGluArgLeuLysThr	957
QY	1176	GGGGAATGTGAAAACCTCCAGTCTCATGCGGAAGTGGAGAGTCTGCGCAAGCGCGCTT	1235
Db	958	SerThrSerAsnValGluGlnTyrGlnAlaMetValThrSerLeuGluGluSerLeuAsn	977
QY	1236	GAGATCGAGGGCAAGGATGAAGAGATCACGAAGACC--GAGGCCAGTGCCTGGAG--	1289
Db	978	LysGluLysGlnValThrGluGluValArgLysAsnIleGluValArgLeuLysGluSer	997
QY	1290	-----CTGAAGAAGAAGTCCAAGAGGAACAACACCACAGCAAGGAA	1331
Db	998	AlaGluPheGlnThrGlnLeuGluLysLysLeuMetGluValGluLysGluLysGlnGlu	1017
QY	1332	CTT-----AGACTAGAAGTGGAAAGCTGCAGAGAGGATGTCTGAGCTGGAG	1379
Db	1018	LeuGlnAspLysArgAlaIleGluSerMetGluGlnGlnLeuSerGluLeuLys	1037
QY	1380	AAG-----CTGGAGGAAGCGTTACGCCGG-----	1403
Db	1038	LysThrLeuSerSerValGlnAsnGluValGlnGluAlaLeuGlnArgAlaSerThrAla	1057
QY	1404	-----AGTAAGTCGGAATGC---ACCCAGCTCCATCTGACCTGGAG	1442
Db	1058	LeuSerAsnGluGlnAlaArgArgAspCysGlnGluGlnAlaLysIleAlaValGlu	1077
QY	1443	AAGGAGAAGAACCTAACCAAAAGACCTGCTG-----AACGAGCTGGAG	1484
Db	1078	AlaGlnAsnLysTyrGluArgGluLeuMetLeuHisAlaAlaAspValGluAlaLeuGln	1097

[illegible]

Db 1454 LysValMetGluThrSerAlaGlnSerSerGlyAspHisGln-----GluGlnHisVal 1471
QY 2166 GCCGTGAGCCAGGAAGCCGAACCTCGACACAGGTTTCGGCTGGAGGAGGCTAAAGTCGT 2225
Db 1472 SerValGlnGluMetGlnGluLeuLysGluThrLeuAsnGlnAlaGluThrLysSerLys 1491
QY 2226 GATTTCAGGCCGAGGTGCAG-----GCTCTCAAGGAGGAAGATCCACGAGCTG 2273
Db 1492 SerLeuGluSerGlnValGluAsnLeuGlnLysThrLeuSerGluLysGluThrGluAla 1511
QY 2274 ATGAACAAGGAAGACCAGCTGTCTCAGCTCCAAGTCCACTATTCCGTCTTCAGCAAGA 2333
Db 1512 ArgAsnLeuGlnGluGlnThrValGlnLeuGlnSerGlnLeuSerArgLeuArgGlnAsp 1531
QY 2334 TTTATGGAAGAAACTAAGAAACAAGAACATGGGAGGAGGTCTCTCAATCTGACCAAG 2393
Db 1532 LeuGlnAspArgThrThrGlnGluGluGlnLeuArgGlnGlnIle-----ThrGlu 1548
QY 2394 GAGCTAGAGCTTTCCAAAGCGCTACAGCCGAGCTCTCAGCCCGAGTGGGAACGCGCAAGG 2453
Db 1549 LysGluGluLysThrArgLys----- 1555
QY 2454 ATGGTGGACGTGCTGGCTCCACTGGGTGCAGACCGAGCGGTGTGCGGGGATGCT 2513
Db 1556 -----AlaIleValAlaAlaLysSerLysIleAlaHisLeuAlaGlyValLysAspGln 1573
QY 2514 GCGGAGGAGGAGACCCCGCTGTGTTTCATTTCGAAATCCTTCCAGGAGGAAATCACATC 2573
Db 1574 LeuThrLysGluAsnGluLeuLysGlnArgAsnGlyAlaLeuAspGlnLysAsp 1593
QY 2574 ATGAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCATGGAA---CGGTCTCGGTCTCTC 2630
Db 1594 GluLeuAspValArgIleThrAlaLeuLysSerGlnTyrGluGlyArgIleSerArgLeu 1613
QY 2631 GACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTGGATTCTCTTGATG 2690
Db 1614 GluArg-----GluLeuArgGluHisGlnGluArgHisLeu 1625
QY 2691 AGAAAAGAGAAAACGGTCTCTCCACTCCGCGAGGAGAAAGGCCAGGCCAAACAGGGGT 2750
Db 1626 GluGlnArgAspGlu---ProGlnGluProSerAsnLys----- 1637
QY 2751 GCAGGGCACCCGGGAGCTGGTCTTAGCACCAAGCAGGGCCAGCCCTACACATCCGT 2810
Db 1638 -----ValProGluGlnGlnArgGlnIleThrLeuLys 1648
QY 2811 GTGACACCA-----GATCATGAGAAACAGCACTGCCACC 2843
Db 1649 ThrThrProAlaSerGlyGluArgGlyIleAlaSerThrSerAspProProThrAlaAsn 1668
QY 2844 CTGGAGATCACAAAGCCCAACATCTGAAGAGTTTTTCTCTAGTACCACCGTCACTCCTACC 2903
Db 1669 IleLysProThrProValValSerThrPro-----SerLysValThrAlaAlaAlaMet 1686
QY 2904 TTAGGCAACACAGAAA---CCAAGAATAACCATTTATCCATCACCCCAATGTCTGTCGCAA 2960
Db 1687 AlaGlyAsnLysSerThrProArgAlaSerIleArgPro-----MetValThr 1702
QY 2961 AAGCCCCAAAAGTGCAGATCCTACTCTCGGCCCCAGAACGAGCCATGTCCCTGTACAGATT 3020
Db 1703 ProAlaThrValThrAsnProThrThrThrThrProThrAlaThrValMetProThrThrGln 1722
QY 3021 ACTACTATTTCCAGAGAGAAGAGCCCGGAAGGTGGAAGGAGCGCCTTTGCCGACAGGCCT 3080
Db 1723 ValGluSerGlnGluAlaMetGlnSerGluGly----- 1733
QY 3081 GCATCCCCCATCCAAATCATGACGGTGTCAACATCTGCAGCTCCCACTGAAATCGCTGTC 3140
Db 1734 -----ProValGluHisValProValPheGlySerThrSerGlySerValArgSerThr 1751
QY 3141 TCTCCTGAATCTCAGGAAGTGCCTATGCGGAAGGACTATCTCTCAAAGTCAACCCCGGAAAAA 3200

Db 1752 SerProAsnValGlnPro---SerIleSerGlnProIleLeuThrValGlnGlnThr 1770
QY 3201 CAAACTGTTCAGCCCCCGTGCAGGAAGTACAATCCAAATGCTAATATCATCACACGGAA 3260
Db 1771 Gln-----AlaThrAlaPheValGlnProThrGln 1780
QY 3261 GACAATAAAATTACATTCACATTACCTGGGTTCTCAGTTTAAAGCGATCTCCTGGCGCTGCCGCT 3320
Db 1781 GlnSerHisProGlnIleGluProAlaAsnGln-----GluLeuSer 1794
QY 3321 GAAGGCGTGAGCCAGTTATCACCGTCCGCGCTGTCAACGTGACAGCGGAGAAAGAGGTT 3380
Db 1795 SerAsnIleValGluValGlnSerSerProValGluArgProSerThrSerThrAla 1814
QY 3381 TCTACAGGCACAGTCTCTCGCTCTCCAGGAACCACTCTCTTCAAGACCCCGTGTAGC 3440
Db 1815 ValPheGlyThrValSerAlaThrProSerSerSerLeuProLysArgThrArgGluGlu 1834
QY 3441 AAAGTGACCAGCACTATA----- 3458
Db 1835 GluGluAspSerThrIleGluAlaSerAspGlnValSerAspThrValGluMetPro 1854
QY 3459 -----ACTATAACCCCGGTC---ACAACGTCTATCCACAGGAAACC 3497
Db 1855 LeuProLysLysLeuLysSerValThrProValGlyThrGluGluValMetAlaGlu 1874
QY 3498 CAATCAGTGTGAGCAAGATGGGTCTCTCAGCGGCTACCCCGCATTCCTATG 3557
Db 1875 GluSerThrAspGlyGluValGluThrGlnValTyrAsnGlnAspSerGlnAspSerIle 1894
QY 3558 TCAAAAGGTATGAAAGCTGGAAGCCAGTAGTGGCAGCCTCAGGAGCAGGAATCTGACC 3617
Db 1895 GlyGluGlyValThr-----GlnGlyAspTyrThr 1904
QY 3618 AAATTCAGCCTCGAGCTGAGACT---CAGTCTATGAAATAGAG-----CTGAAG 3665
Db 1905 ProMetGluAspSerGluGluThrSerGlnSerLeuGlnIleAspLeuGlyProLeuGln 1924
QY 3666 AAATCTGCAGCCAGCAGCACTGCCTCTCTTGGAGGG---GGGAAGGGC 3710
Db 1925 SerAspGlnGlnThrThrThrSerSerGlnAspGlyGlnGlyLysGly 1940

RESULT 3
S28261
centromere protein E - human
N;Alternate names: centromere 312K protein; kinesin-related protein CENP-E
C;Species: Homo sapiens (man)
C;Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C;Accession: S28261
R;Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
Nature 359, 536-539, 1992
A;Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
A;Reference number: S28261; MUID:93024922; PMID:1406971
A;Accession: S28261
A;Molecule type: mRNA
A;Residues: 1-2663 <YEN>
A;Cross-references: EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865
C;Genetics:
A;Gene: GDB:CENPE
A;Cross-references: GDB:361164; OMIM:117143
A;Map position: 4q24-4q25
C;Superfamily: centromere protein E; kinesin motor domain homology
C;Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
F;7-335/Domain: kinesin motor domain homology <KMOT>
F;86-93/Region: nucleotide-binding motif A (P-loop)
F;486-2183/Domain: coiled coil #status predicted <COI>
F;92/Binding site: ATP (Lys) #status predicted

Alignment Scores:
Pred. No.: 1.19e-14 Length: 2663
Score: 473.50 Matches: 289
Percent Similarity: 39.37% Conservative: 263
Best Local Similarity: 20.61% Mismatches: 461

Query Match: DB:	6.12% 1	Indels: Gaps:	389 64
US-10-788-793-1 (1-4364) x S28261 (1-2663)			
QY	171	TCAGAAAGATGCAAAAAAGAACAAAGGCCAATCGGAAGGAGGAGGAT-----	215
Db	1382	SerGlnSerLysGlnGluGlnSerLeuAsnMetLysGluLysAspAsnGluThrThrLys	1401
QY	216	-----GTCATGGCTTCCGGAACTATCAAAGGCACCTC	248
Db	1402	IleValSerGluMetGluGlnPheLysProLysAspSerAlaLeuLeuArgIleGluIle	1421
QY	249	AAACCATCTGGAGAAAGTGAGAAAAAGACTAAG-----	293
Db	1422	GluMetLeuGlyLeuSerLysArgLeuGlnGluSerHisaspGluMetLysSerValAla	1441
QY	294	TTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGGAGTTG---	350
Db	1442	LysGluLysaspLeuGlnArgLeuGlnGluValLeuGlnSerGluSerAspGlnLeu	1461
QY	351	CGAGAAGATGTCATCCATGCTGAGGACAGAG---AAACCAAGCCCCGAGGTTCTGGAG	407
Db	1462	LysGluAsnIleLysGluIleValAlaLysHisLeuGluThrGluGluLeuLysVal	1481
QY	408	GCACACTATGATCTGCAGAACCTGAGAAAGTGCTTCGGGTCTCTGCACCGAGATGCCATC	467
Db	1482	AlaHisCysCysLeuLysGluGlnGluGluThrIleAsnGluLeuArg-----Val	1498
QY	468	CTTGCTCAAGAGAAGTCCATPAGGAGAAGACGCTCTATGAGAAACCTATCTCAGAGCTG---	524
Db	1499	AsnLeuSerGluLysGluThrGluIleSerThrIleGlnLysGlnLeuGluAlaIleAsn	1518
QY	525	GACAGACTGGAGGAAAAGCAGAAGGAGAGCTACCGCCGCATGCTAGAGAGCTGCTGTCG	584
Db	1519	AspLysLeuGlnAsnLysIleGlnGluIleTyrGluLysGlu---GluGlnLeuAsnIle	1537
QY	585	GCTGAG-----	602
Db	1538	LysGlnIleSerGluValGlnGluAsnValAsnGluLeuLysGlnPheLysGluHisArg	1557
QY	603	CGC-----ACCGTGACGAGCTGGAGAACGAGAAGCACAAGCACACTGACTAC	650
Db	1558	LysAlaLysAspSerAlaLeuGlnSerIleGluSerLysMetLeuGluLeuThrAsnArg	1577
QY	651	ATGAACAAGACGACGACTTCACCAACCTGCTGGAGCGAGGCGAGAGAGGTTGAAAAAG	710
Db	1578	LeuGlnGluSerGlnGluGluIleGlnIleMetIleLysGluLysGluMetLysArg	1597
QY	711	CTCCTTGAAACAAGAAAAGCTTACCAAGCCCCGAAAAGAA-----AAGGAAAAACGCTPAAG	764
Db	1598	Val-----GlnGluAlaLeuGlnIleGluArgaspGlnLeuLysGluAsnThrLys	1614
QY	765	CGGCTC---AACAACTTCGAGATGAGCTTGTGAAG-----CTCAAGTCCTTC	809
Db	1615	GluIleValAlaLysMetLysGluSerGlnGluLysGluTyrGlnPheLeuLysMetThr	1634
QY	810	GCCCTCATGTTGGTGACGAGAGGCAGATGCACATCGAGCAACTG-----GGCCTGCAG	863
Db	1635	AlaValAsnGluThrGlnGluLysMetCysGluIleGluHisLeuLysGluGlnPheGlu	1654
QY	864	AGTCAGAAAGTCCAG-----GACCTCACTCAGAAAGCTG	896
Db	1655	ThrGlnLysLeuAsnLeuGluAsnIleGluThrGluAsnIleArgLeuThrGlnIleLeu	1674
QY	897	AGGGAGGAGGAAGAAAAACTCAAAGCGGTCACTTTACAAATCCAAAGGAAGACCGC-----	950
Db	1675	HisGluAsnLeuGluMetArgSerValThr---LysGluArgaspLeuArgSer	1693
QY	951	---CAGAAGCTGCTCAAGTTAGAAGTGGAC-----TTCGAACACAAAGGCC	992
Db	1694	ValGluGluThrLeuLysValGluArgaspGlnLeuLysGluAsnLeuArgGluThrIle	1713

QY	993	TCGAGGTTTCCCAGGACACGAAGAGATGAACGCCAAATTGGCGAATCAAGAATCTCAC	1052
Db	1714	ThrArgaspLeuGluLysGlnGluGluLeuLysIleValHisMetHisLeuLysGluHis	1733
QY	1053	AACCGGCAACTTCGACTCAAACTGGTTGGCTTA---TCGCAAAAGGATTGAGGAGCTGGAA	1109
Db	1734	GlnGluThrIleAsp---LysLeuArgGlyIleValSerGluLysThrAsnGluIleSer	1752
QY	1110	GAGACCAATAAAGCCTTCAGAAGGCAGAGGAAGAGCTCCAG-----GAGCTG	1157
Db	1753	AsnMetGlnLysAspLeuGluHisSerAsnAspAlaLeuLysAlaGlnAspLeuLysIle	1772
QY	1158	AGAGAGAAAATTGCCAAAGGGGAATGTGGAAACTCCAGTCTCATGGCGGAAGTGGAGAGT	1217
Db	1773	GlnGluGluLeuArgIleAlaHisMetHisLeuLysGluGlnGlnGluThrIleAspLys	1792
QY	1218	CTCGCAAGCGCGTGTGATGGAGGGCAAG-----GATGAAGAGATCACG	1265
Db	1793	LeuArgGlyIleValSerGluLysThrAspLysLeuSerAsnMetGlnLysAspLeuGlu	1812
QY	1266	AAGACGAGGCCAGTGCCGGGAGCTGAAGAAGAGCTCCAAGAGGAACAACACCACAGC	1325
Db	1813	AsnSerAsnAlaLysLeuGlnGluLysIleGlnGluLeuLysAlaAsnGluHisGlnLeu	1832
QY	1326	AAGGAACCTAGACTAGAAGTGGAGAAGCTGCAGAAGAGGATGTCTGAGCTGGAGAAGCTG	1385
Db	1833	IleThrLeuLysLysaspValAsnGluThrGlnLysValSerGluMetGluGlnLeu	1852
QY	1386	-----GAGGAAGCGTTTCAGCCGCGGAGTAAGTCGGAATGTCACCCAGCTCCAT	1430
Db	1853	LysLysGlnIleLysAspGlnSerLeuThrLeuSerLysLeuGluIleGluAsnLeu---	1871
QY	1431	CTGAACCTGGAGAAGAGAAACCTTAACCAAGACCTGCTGAACGAGCTGGAGGTGGTC	1490
Db	1872	-----AsnLeuAlaGlnGluLeuHisGluAsnLeuGluGluMet	1884
QY	1491	AAGACTCGAGTTAAAGAA-----CTCGAATGCTCCGAGAGTAGACTGGAGAAGGCC	1541
Db	1885	LysSerValMetLysGluArgAspAsnLeuArgValGluGluThrLeuLysLeuGlu	1904
QY	1542	GAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAGTCTTCACTGTGTGATGCTGTGGAT	1601
Db	1905	ArgaspGlnLeuLysGluSerLeuGlnGluThrLysAlaArgaspLeuGluIleGlnGln	1924
QY	1602	GAGAGGAAAAATATGATGGAGAAAAATAAAGCAAGAGAGGAAAGTGGATGGGTTGAAT	1661
Db	1925	GluLeuLysThrAlaArgMetLeuSerLysGluHisLysGluThrValaspLysLeuArg	1944
QY	1662	AAAAACTTTAAGGTGGAGCAGGGAAGAAAGTCATGGATGTGACGGGAAAAGCTAATCGAGGAA	1721
Db	1945	Glu-----LysIleSerGluLysThrIleGlnIleSeraspIle	1957
QY	1722	AGCAAGAAGCTTTTAAAACTCAAATCTGAAATCGAGGAAAAGGAGTACAGTCTGACAAAAG	1781
Db	1958	GlnLysAspLeuAspLysSerLysAspGluLeuGlnLysLysIleGlnGluLeuGlnLys	1977
QY	1782	GAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGGAAGAAAGAGTCCCTGTGAACGTGAGC	1841
Db	1978	LysGluLeuGlnLeu-----LeuArgValLysGluAsp-----	1988
QY	1842	TGCAGGTGACTTACTAAAGACGGCTTGATGGCATAGAGAGGTAGAAAGGAAATA	1901
Db	1989	-----ValAsnMetSerHisLysLysIleAsnGluMetGluGlnLeuLysLysGlnPhe	2006
QY	1902	AACCGAGGTAGTCTGTCGAAGGGGTCTGAGTTTCACCTGCCCCGGAAGACAATAAGATCAGA	1961
Db	2007	GluProAsnTyrLeuCysLys-----CysGluMetAspAsn-----Phe	2019
QY	1962	GAACCTAACGCTTGAAATCGAGAGACTGAAGAAAACGGCTCCAGCAGTTGGAGTGGTGGAG	2021
Db	2020	GlnLeuThr-----LysLysLeuHisGluSerLeuGluGluIleArgIleVal---	2035
QY	2022	GGGGACTTGATGAAGACCGGAGACCGAATATGACCAGTTGGAGCAGAAGTTTCAGAACCCGAG	2081

Db 2036 -----AlaLysGluArgAspGluLeuArgArgIleLysGluSerLeuLysMetGlu 2052
QY 2082 CAGGATAAGGCAGAACTTCTCTCCAGCAGCTCGAGGAATC-----AAA 2126
Db 2053 ArgAspGln-----PheIleAlaThrLeuArgGluMetIleAlaArgAspArgGlnAsn 2070
QY 2127 CACCAAAATGGCCAAGCACAAAGCCATA----- 2153
Db 2071 HisGlnValLysProGluLysArgGluLeuSerAspGlyGlnGlnHisLeuMetGluSer 2090
QY 2154 -----GAGAAA-----GGGGAGGCCGTGAGCCAGGAAGCCGAACTGCGACAC 2195
Db 2091 LeuArgGluLysCysSerArgIleLysGluLeuLeuLysArgTyrSerGluMetAspAsp 2110
QY 2196 AGGTTTCGGCTGGAGGAGGCTAAAGTCGTGATTACAGGCCGAGGTGCAG----- 2246
Db 2111 HisTyrGluCysLeuAsnArgLeuSerLeuAspLeuGluLysGluIleGluPheHisArg 2130
QY 2247 -----GCTCTCAAGGAGAAGATCCAC 2267
Db 2131 IleMetLysLysLeuLysTyrValLeuSerTyrValThrLysIleLysGluGlnHis 2150
QY 2268 GAGCTGATGAACAAGGAGCAGCTGTCTCAGCTCCAACTCGACTATTCCGGTCTCTCAG 2327
Db 2151 GluCysIleAsnLys-----PheGlu 2157
QY 2328 CAAAGATTATGGAAGAAGAACTAAGAACAAAGACATGGGGAGGGAGGTCCTCAATCTG 2387
Db 2158 MetAspPheIleAspGluValGluLysGlnLysGluLeuLeuIleLysIleGlnHisLeu 2177
QY 2388 ACCAAGGAGCTAGAGCTTTCCAAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGC 2447
Db 2178 GlnGlnAspCysAspValProSerArgGluLeuArgAspLeuLysLeuAsnGlnAsn-- 2196
QY 2448 CGAAGGATGGTGAGCTGCGCTGTGGCTCCACTGGGGTGCAGACCAGGCGGTGTGCGGG 2507
Db 2197 -----MetAspLeuHisIle-----GluGluIleLeuLys 2206
QY 2508 GATGCTGGGAGGAGGAGACCCCGGCTGTGTTTCATTCGCAAACTCTTCCAGGAGGAAAAT 2567
Db 2207 AspPheSerGluSerGluPheProSer-----IleLysThrGluPheGlnGln----- 2222
QY 2568 CACATCATGAGTAATCTTCACAGGTA----- 2594
Db 2223 ---ValLeuSerAsnArgLysGluMetThrGlnPheLeuGluGluTrpLeuAsnThrArg 2241
QY 2595 -----GGCCTGAAGAAACCCATGGAAACGGTCTCGTCTCGTCTC 2630
Db 2242 PheAspIleGluLysLeuLysAsnGlyIleGlnLysGluAsnAspArgIleCysGlnVal 2261
QY 2631 -----GACAGGTATCCCCCAGCAGCGAATGAG----- 2657
Db 2262 AsnAsnPhePheAsnAsnArgIleIleAlaIleMetAsnGluSerThrGluPheGluGlu 2281
QY 2658 -----CTCACCATGAGGAAGTCTTGGATTCTTGGATGAGAAAAAGAGAAACCGTCTCT 2711
Db 2282 ArgSerAlaThrIleSerLysGluTrp-----GluGlnAspLeu 2294
QY 2712 TCCACTCCGAGGAGAAAGGCCCCAGGCCAACCCAGGGTGCAGGGCACCCCGGGGAGCTG 2771
Db 2295 LysSerLeuLysGluLysAsnGluLys----- 2303
QY 2772 GTCCTAGCACCAAGAGCGCCAGCCCTACACATCCGTGTGACACCAGATCATGAGAAC 2831
Db 2304 -----LeuPheLysAsnTyrGlnThrLeu-----LysThr 2313
QY 2832 AGCACTGCCACCCCTGGAGATCACAAAGCCCCACATCTGAAGAGTTTTTCTCTAGTACCACC 2891
Db 2314 SerLeuAlaSerGlyAlaGlnValAsnProThrThrGlnAsp----- 2327
QY 2892 GTCATTCTTACCTTAGGCAACCAAGAAACCAAGAAATACCATTATTCCATCACCACCAATGTC 2951
||||| :||| :||| :|||

Db 2328 -----AsnLysAsnProHisValThr---SerArgAlaThrGlnLeu 2340
QY 2952 ATGTGCGAAAGCCCAAAAGTGCAGATCCTACTCTC-----GGCCCAGAACGAGCCATG 3005
Db 2341 ThrThrGluLysIleArgGluLeuGluAsnSerLeuHisGluAlaLysGluSerAlaMet 2360
QY 3006 TCC-----CCTGTCAACGATTACTACTATT 3029
Db 2361 HisLysGluSerLysIleIleLysMetGlnLysGluLeuGluValThrAsnAspIleIle 2380
QY 3030 TCCAGAGAGAAGAGCCCGGAAGGTGGAAGAGCGCCTTTCCGACAGGCCCTGCATCCCCC 3089
Db 2381 AlaLysLeuGlnAlaLysValHisGluSerAsnLysCysLeuGluLysThrLysGluThr 2400
QY 3090 ATCCAAATCATGACGGTGTCAACATCTGCAGCTCCCACTGAAATCGCTGTCTCTCTGAA 3149
Db 2401 IleGlnValLeuGlnAspLysValAla-----LeuGlyAlaLysProTyr 2415
QY 3150 TCTCAGGAAGTG---CCTATGGGAAGGACTATCTCAAAAGTCACCCCGGAAAAACAACACT 3206
Db 2416 LysGluGluIleGluAspLeuLysMetLysLeuValLysIleAspLeuGluLysMetLys 2435
QY 3207 GTTCCAGCCCCCGTGCAGAGTACAACCTCCAATGTAATATCATCACCGGAGACAAAT 3266
Db 2436 AsnAlaLysGluPheGluLys---GluIleSerAlaThrLysAlaThrValGluTyrGln 2454
QY 3267 AAAATTACATTCACCTG-----GGTTCTCAGTTTAAAGCGATCTCCT 3308
Db 2455 LysGluValIleArgLeuLeuArgGluAsnLeuArgArgSerGlnGlnAlaGlnAspThr 2474
QY 3309 GGGCTGCGCGTGAAGCGTGAGCCCGAGTTATCACCGTCCGGCGCTGTC----- 3356
Db 2475 SerValIleSerGluHisThrAspProGlnProSerAsnLysProLeuThrCysGlyGly 2494
QY 3357 -----AACGTGACAGCG 3368
Db 2495 GlySerGlyIleValGlnAsnThrLysAlaLeuIleLeuLysSerGluHisIleArgLeu 2514
QY 3369 GAGAAGGAGGTTTCT-----ACAGGCACAGTC 3395
Db 2515 GluLysGluIleSerLysLeuLysGlnGlnAsnGluGlnLeuIleLysGlnLysAsnGlu 2534
QY 3396 CTTCGCTCTCCCAGGAACCACTCTCTTCAAGA----- 3428
Db 2535 LeuLeuSerAsnAsnGlnHisLeuSerAsnGluValLysThrTrpLysGluArgThrLeu 2554
QY 3429 -----CCCGGTGCTAGCAAAAGTGACC 3449
Db 2555 LysArgGluAlaHisLysGlnValThrCysGluAsnSerProLysSerProLysValThr 2574
QY 3450 AGCACTATAACT-----ATAACCCCGGTCAAA-----CGTCAT----- 3483
Db 2575 GlyThrAlaSerLysLysGlnIleThrPro-SerGlnCysLysGluArgAsnLeuG1 2594
QY 3484 -----CCACACGAGGAACCCAATCAGTGT---CAGGACAAGATGGGTCTCTCAG 3530
Db 2594 nAspProValProLysGluSerProLysSerCysPheAspSerArgSerLys-SerL 2614
QY 3531 CGGCCTACCCCCACCGCATTCCTATGTCAAAAGGTATGAA----- 3571
Db 2614 euProSerProHisProValArgTyrPheAspAsnSerSerLeuGlyLeuCysProGluV 2634
QY 3572 -----AGCTGGAAGCCAGTAGTG 3590
Db 2634 alGlnAsnAlaGlyAlaGluSerValAspSerGlnProGlyProTrpHisAlaSerSerG 2654
QY 3591 GC 3592
|||
Db 2654 ly 2654
RESULT 4
B43402
myosin heavy chain-B, neuronal - chicken

N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C;Accession: B43402; A43402
R;Takahashi, M.; Kawamoto, S.; Adelstein, R.S.
J. Biol. Chem. 267, 17864-17871, 1992
A;Title: Evidence for inserted sequences in the head region of nonmuscle myosin specific yosin.

A;Reference number: A43402; MUID:92388144; PMID:1355479
A;Accession: B43402
A;Molecule type: mRNA
A;Residues: 1-2007 <TAK>
A;Cross-references: GB:M93676; NID:g212448; PIDN:AAA48988.1; PID:g212452
A;Note: the sequence of residues 212-221 and 632-652 and the corresponding nucleotide se
A;Accession: A43402
A;Molecule type: mRNA
A;Residues: 1-211;222-631;653-2007 <TA2>
A;Cross-references: GB:M93676; NID:g212448; PIDN:AAA48988.1; PID:g212449
A;Note: sequence extracted from NCBI backbone (NCBIN:112864)
C;Comment: Alternatively spliced segments 1 and 2 are found exclusively in nonmuscle myo
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
F;1-2007/Product: myosin heavy chain-B, neuronal #status predicted <MYN>
F;1-211,222-631,653-2007/Product: myosin heavy chain-B, nonmuscle #status predicted <MYN>
F;88-802/Domain: myosin motor domain homology <MMOT>
F;178-185/Region: nucleotide-binding motif A (P-loop)
F;212-221/Region: alternatively spliced segment 1 #status experimental
F;559-593/Region: actin binding #status predicted
F;632-652/Region: alternatively spliced segment 2 #status experimental
F;692-714/Region: actin binding #status predicted
F;875-2007/Domain: coiled coil #status predicted <COI>
F;875-1315/Region: S2

F;1316-2007/Region: light meromyosin
F;129/Modified site: N6,N6,trimethyllysine (Lys) #status predicted
F;184/Binding site: ATP (Lys) #status predicted
F;732,742/Active site: Cys #status predicted
F;1954/Binding site: phosphate (Thr) (covalent) #status predicted
F;1987/Binding site: phosphate (Ser) (covalent) #status predicted

Alignment Scores:
Pred. No.: 1.43e-14 Length: 2007
Score: 471.50 Matches: 270
Percent Similarity: 40.82% Conservative: 257
Best Local Similarity: 20.91% Mismatches: 488
Query Match: 6.10% Indels: 276
DB: 1 Gaps: 51

US-10-788-793-1 (1-4364) x B43402 (1-2007)

Qy	101	AAGTTCACTAACGGGCATGTCTCTCTGCCCCAAGTCTCCATCATCAGCAGTGATGTGG	160
Db	848	LysIleLeuGlnArgAsnCysAlaAlaTyrLeuLysLeuArgHisTrpGln---TrpTrp	866
Qy	161	TAAGGGCCCTCAGAAAGATGCAAAAAAGAACAAAGGCCCAATCGGAAGGAGGAGGATGTCAT	220
Db	867	-ArgValPheThrLysValLysProLeuLeuGlnValThrArgGlnGluGluLeuGl	886
Qy	221	GGCTTCGGAACTATCAAAAGGCACCTCAAAACCATCTCGAGAGAAAGTGAGAAAAAGACTAA	280
Db	886	nAlaLysAspGluLeuLeuMetLysValLys-----GluLysGlnThrLy	901
Qy	281	GAAGTCTGTGGAGTTATCCAAG-----GAGGACCTCATCCAGCTCCTGAG	325
Db	901	sValGluAlaGluLeuGluGluMetGluArgLysHisGlnGlnLeuLeuGluLysAs	921
Qy	326	TATCATGGAAGGGGAGTTGCAGGGCTCGAGAAAGATGTCCATCCACATGCTGAGGACAGAGAA	385
Db	921	nIleLeuAlaGluGlnLeuGlnAlaGluThrGluLeuPhe-----	934
Qy	386	AACCAAGCCCGAGGTTCTGGAGGCACACTATGGATCTGCAGAACCTGCAGAAAGTGCTCG	445
Db	935	-AlaGluAlaGluGluMetArgAlaArgLeu---AlaAlaLysLysGlnGluLeuGl	953

Qy	446	GGTCCTGCACCGAGATGCCATCCTTGCTCAAGAGAAAGTCCATAGGAGAAAGACGCTCTATGA	505
Db	953	uIleLeuHis-----AspLeuGluSerArgValGluGlu-----Gl	965
Qy	506	GA AACCTATCTCAGAGCTGGACAGACTGGAGGAAAAAGAGAGAGACGTACCGCCGCAT	565
Db	965	uGluGluArgAsnGlnIleLeuGlnAsnGluLysLysMetGlnGlyHisIleGlnAs	985
Qy	566	GCTAGAGCAGCTGCTGCTGGCTGAGAAAGTGTACAGGGCGCACCGTGTACGAGCTGGAGAA	625
Db	985	pLeuGluGluGlnLeuAspGluGluGlyAlaArgGln-----LysLeuGlnLe	1002
Qy	626	CGAGAAGCACAAGCACACTGACTACATGAACAAGAGCGCAGCAGCTTCACCAACCTGCTGGA	685
Db	1002	uGluLysValThrAlaGluAlaLysIleLysLysMetGluGluGluIleLeuLeuGl	1022
Qy	686	GCAGGACGAGAGAGGTTG-----AAAAAGCTCCTTGAACAA-----GAAAA	727
Db	1022	uAspGlnAsnSerLysPheLeuLysGluLysLysLeuMetGluAspArgIleAlaGluCy	1042
Qy	728	AGCTTACCAAGCCCCGCAAGAAAAAGGAAAAACGCTAAGCGGCTCAACAAACTTCGAGATGA	787
Db	1042	sThrSerGlnLeuAlaGluGluGluLysAlaLysAsnLeuAlaLysLeuLysAsnLy	1062
Qy	788	GCTTGTGAAGCTCAAGTCCTTCGCCCTCATGTTGGTGACGAGAGGCAGATGCACATCGA	847
Db	1062	s-----	1062
Qy	848	GCAACTGGGCCTGCAGAGTCAGAAAAGTCCAGGACCTCACTCAGAAAGCTGAGGGAGGAGA	907
Db	1063	-----GlnGluMetMetIleThrAspLeuGluGluArgLeuLysLysGluGl	1078
Qy	908	A-----GAAAACTCAAAAGCGGTCACCTTACAAATCAAGGAAGACCG	949
Db	1078	uLysThrArgGlnGluLeuGluLysAlaLysArgLysLeuAspGlyGluThrThrAspLe	1098
Qy	950	CCAG---AAGCTGCTCAAGTTAGAGTGGACTTCGAAACAACAAGGCCTCGAGGTTTCCCA	1006
Db	1098	uGlnAspGlnIleAlaGluLeuGlnAlaGlnIleGluGluLeuLysIleGlnLeuAlaLy	1118
Qy	1007	GGAGCACGAAGAGATGAACGCCCAATTTGGCG-----AATCAAGAATCTCACAAACCGGCA	1060
Db	1118	sLysGluGluGluLeuGlnAlaAlaLeuAlaArgGlyAspGluGluAlaValGlnLysAs	1138
Qy	1061	ACTTCGACTCAAACTGGTT-----GGCTTATCGCAAGAGATTGAGGAGCT	1105
Db	1138	nAsnAlaLeuLysValIleArgGluLeuGlnAlaGlnIleAlaGluLeuGlnGluAspLe	1158
Qy	1106	GGAAGACACCAATAAAAGCCTTCAGAAGGCAGAGAAAGAGCTCCAGGAGCTGAGAGAGAA	1165
Db	1158	uGluSerGluLysAlaSerArgAsnLysAlaGluLysGlnLysArgAspLeuSerGluGl	1178
Qy	1166	AATTGCCAAAGGGGAATGTGGAAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCCGCAA	1225
Db	1178	uLeu-----GluAlaLeuLysThrGluLeuGluAsp-----	1188
Qy	1226	GCGCTGCTTGAGATGGAGGGCAAGGATGAAGAGATC---ACGAAGACCGAGGCCAGTG	1282
Db	1189	-----ThrLeuAspThrThrAlaAlaGlnGlnGluLeuArgThrLysArgGluGlnGluVa	1207
Qy	1283	CCGGAGCTGAAGAAGAGCTCCAGAGAGAA---GAACACCACAGCAAGGAACCTTAGACT	1339
Db	1207	lAlaGluLeuLysLysAlaIleGluGluGluThrLysAsnHisGluAlaGlnIleGln--	1226
Qy	1340	AGAAGTGGAGAGCTGCAGAAGAGGATGTCTGAGCTGGAGAGCTGGAGGAGCGTTTCAG	1399
Db	1227	-----GluIleArgGlnArgHisAlaThrAlaLeuGluGluLeuSerGluGlnLeuGl	1244
Qy	1400	CCGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAGGAGAAG-----	1451
Db	1244	uGlnAlaLys-----ArgPheLysAlaAsnLeuGluLysAsnLysGlnGlyLe	1260
Qy	1452	---AACCTAACCAAAAGACCTGCTGAACGAGCTGGAGGTGGTCAAG-----	1493

Db	1260	uGluSerAspAsnLysGluLeuAlaCysGluValLysValLeuGlnValLysAlaG1	1285
QY	1494	-----AGTCGAGTTAAAGAACTCGAATGCTCCGAGAG	1525
Db	1280	uSerGluHisLysArgLysLysLeuAspAlaGlnValGlnGluLeuThrAlaLysValTh	1300
QY	1526	TAGACTGGAGAAGGCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAAGTCCCTCAC	1585
Db	1300	rGluGlyGluArgLeuArgValGluLeuAlaGluLysAlaAsnLysLeuGln-----	1317
QY	1586	TGTGATGCTGTTGATGAGAGGAAAAATATGATGGAGAAAAATAAAGCAAGAGAGGAA	1645
Db	1318	-----AsnGluLeuAspAsnValSerSerLeuLeuGluAlaGluLysLeuGln-----	1333
QY	1646	AGTGGATGGTTGAATAAAAACTTTAAGGTGGAGCAGGGAAGATCATGGATGTGACGGA	1705
Db	1333	sGlyIleLysPheAlaLysAspAlaAlaSerLeuGluSerGlnLeuGlnAspThrGlnG1	1353
QY	1706	AAAGCTAATCGAGGAACGAAGCTTTTAAACCTCAAATCT-----GAAATGGA	1756
Db	1353	uLeuLeuGlnGluThrArgGlnLysLeuAsnLeuSerSerArgIleArgGlnLeuG1	1373
QY	1757	GGAAAGGAGTACAGTCTGACAAAGGAGAGGGATGAGCTGATGGTAACTGAGGAGCGA	1816
Db	1373	uGluGluLysAsnAsnLeuGlnGlnGlnGluGluGluGluAlaArgLysAsnLe	1393
QY	1817	AGAAGAAAGGTCCTGTGAAGCTGAGCTGCAGTGTAGACTTACTAAAGAAGCGGCTTGAT--	1874
Db	1393	uGluLysGlnMetLeuAlaLeuGlnAlaGlnLeuAlaGluAlaLysLysValAspAs	1413
QY	1875	-----GGCATAGAGGAGGTAGAAAGGGAAATAAAACCGAGGTAGGTC	1915
Db	1413	pAspLeuGlyThrIleGluGlyLeuGluGlu-----AsnLysLysLysLe	1428
QY	1916	GTGCAAGGGGTCTGAGTTACCTGCCCGAAGACATAAGATCAGAGAACTAACGCTTGA	1975
Db	1428	uLeuLysAspMetGluSerLeu-----SerGlnArgLeuGluGluLysAlaMetAl	1445
QY	1976	AATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTGGAGGGGACTTGATGAA	2035
Db	1445	aTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAspAspLeuMetVa	1465
QY	2036	GACCGAGGACGAATATGACCAGTTGGAGCAGAAGTTCAGAAACCGAGCAGGATAAGGCAAA	2095
Db	1465	lAspLeuAspHisGlnArgGlnIleValSerAsnLeuGluLysLysGlnLysLys-----	1483
QY	2096	CTTCCTCTCCAGCAGCTCGAGGAAATCAAACACCAAAATGGCCAAAGCAC-----	2144
Db	1484	---PheAspGlnMetLeuAlaGluGluLysAsnIleSerAlaArgTyrAlaGluGluAr	1502
QY	2145	-----AAAGCCATAGAGAAAGGGGAGGC	2167
Db	1502	gAspArgAlaGluAlaGluAlaArgGluLysGluThrLysAlaLeuSerLeuAlaArgAl	1522
QY	2168	CGTGAGCCAGGAAGCCGAAGCTGCGACACAGGTTT-----	2201
Db	1522	aLeuGluGluAlaLeuGluAlaLysGluGluPheGluArgGlnAsnLysGlnLeuArgAl	1542
QY	2202	-----CGGCTGGA	2209
Db	1542	aAspMetGluAspLeuMetSerSerLysAspAspValGlyLysAsnValHisGluLeuG1	1562
QY	2210	GGAGGCTAAAAGTCGTATTACAGGCCGAGGTGAGGCTCTCAAGGAGAGACATCCACGA	2269
Db	1562	uLysSerLys---ArgThrLeuGluGlnGlnValGluGluMetArgThrGlnLeuGluG1	1581
QY	2270	GCTGATGAACAAG-----GAAGACCAGCTGTCTCAGCTCCCAAGTCGACTATTC	2317
Db	1581	uLeuGluAspGluLeuGlnAlaThrGluAspAlaLysLeuArgLeuGluValAsnMetG1	1601
QY	2318	GGTC-----CTTCAGCAAGAGTTTATGGAAGAGAAACTAA	2353

Db	1601	nAlaMetLysAlaGlnPheGluArgAspLeuGlnAlaArgAspGluGlnAsnGluGluLy	1621
QY	2354	GAACAAGAACATGGGGAGGAGGTCCTCAATCTGACCAAGGAGCTAGAGCTTCCAAGCG	2413
Db	1621	sLysArgMetLeuValLysGlnValArgGluLeuGluAlaGluLeuGlu--AspGluAr	1640
QY	2414	CTACAGCCGAGTCTCAGGCCGAGTGGGAACGGCCGAAGGATGGTGACGTCCTGTGGC	2473
Db	1640	gLysGlnArgAlaLeuAlaValAlaLysLysLysMetGluMetAspLeuLysAspLe	1660
QY	2474	CTCCACTGGGTGCAGACCGAGGGGTGTGCGGGGATGTCGCGGAGGAGACCCCGGC	2533
Db	1660	uGluGlyGlnIleGluAlaAlaAsnLysAlaArgAspGluAlaIleLysGlnLeuArgLy	1680
QY	2534	TGTGTTCAATCCG---AAATCTTCCAGGAGAA-----AA	2566
Db	1680	sLeuGlnAlaGlnMetLysAspTyrGlnArgGluLeuGluAlaArgAlaSerArgAs	1700
QY	2567	TCACATCATGAGTAATCTTCGACAGGTAGGCCTGAAG--AAACCCATGGAACGGTCCTC	2623
Db	1700	pGluIlePheAlaGlnSerLysGluSerGluLysLysLysLysGlyLeuGlu-----	1717
QY	2624	GGTCCTCGACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGAAAGTCTTGGATTCC	2683
Db	1718	-----AlaGluIleLeuGlnLeuGlnGluGluPheAlaAl	1729
QY	2684	TTGGATGAGAAAAAGAGAAACGGTCTTCCACTCCGCAGGAGAAAGGCCCAGGCCAAA	2743
Db	1729	aSerGluArgAlaArgHisAlaGluGlnGluArgAspGluLeuAlaAspGluIleAl	1749
QY	2744	CCAGGTGCAGGCCACCCCGGGGAGTGGTCTTAGCACCAAAAGCAGGCCAGCCCTTACA	2803
Db	1749	aAsnSerAlaSerGlyLysSerAlaLeuLeuAspGluLysArgArgLeuGluAlaArgIl	1769
QY	2804	CATCCGTGTCACACCAGATCATGTAGAACAGCAGCTGCCACCCCTGGAGATCACAAGCCCCAC	2863
Db	1769	eAlaGlnLeuGluGluGluLeuGluGluGlnSerAsnMetGluLeu-----Le	1786
QY	2864	ATCTGAAGATTTTCTCTAGTACCAACCGTCATTCTTACCTTAGGCAACCCAGAAACCAAG	2923
Db	1786	uAsnGluArgPheArgLysThrThrLeuGlnValAspThrLeuAsnSerGlu-----	1803
QY	2924	AATAACCATATTCCATCACCCCAATGTCTATGTCGCAAAAGCCCAAGTGCAGATCCTAC	2983
Db	1804	-----LeuAlaGlyGluArgSerAlaAlaGlnLysSerGluAsnAlaArgGlnGl	1820
QY	2984	TCTCGGCCCAGAACGAGCCATGTCCCTGTCCAGGATTACTACTATTCCAGAGAGAAGAG	3043
Db	1820	nLeu-----GluArgGlnAsnLysGluLeuLys-----AlaLysLeuGlnGl	1834
QY	3044	CCCGGAAGTGGA---AGGAGCGCTTTTGCACAGGCCTGCATCCCCCATCCAATCAT	3100
Db	1834	uLeuGluGlySerValLysSerLysPhe-----LysAlaThrIleSe	1848
QY	3101	GACGGTGTCAACATCTGCAGCTCCCACTGAAATCGCTGTCTCTCTGAATCTCAGGAAGT	3160
Db	1848	rThrLeuGluAlaLysIleAlaGlnLeuGluGlnLeuGluGlnAlaLysGluAr	1868
QY	3161	GCCTATGGGAAGGACTATCCTCAAGTCACCCCGGAAAAA---CAAACTGTTCCAGCCCC	3217
Db	1868	gAlaAlaAlaAsnLysLeuValArgArgThrGluLysLysLeuLysGluValPheMetGl	1888
QY	3218	CGTGGGAAGTACAACTCCAATGCTTAATATCATCACCCAGGAAGACAATAAAATTCACAT	3277
Db	1888	nValGluAspGluArgHisAlaAspGlnTyrLysGluGlnMetGluLysAlaAsnAl	1908
QY	3278	TCACCTGGGTCTCAGTTTAAGCGATCTCTGGGCCTGCCGCTGAAGCGTGAGCCCCAGT	3337
Db	1908	aArgMet---LysGlnLeuLysArgGlnLeuGluAlaGluGluAlaThrArgAl	1927
QY	3338	TATCACCTCGGCCTGTCAACCGTGCAGCGGAGAGGAGTTTCTACAGGCACAGTCCT	3397
Db	1927	aAsnAlaSerArgArg-----LysLeuGlnArgGluLeuAspAlaThr-----	1942

QY 3398 TCGCTCTCCAGGAACACCTCTCTTCAAGACCCGGTGTAGCAAAAGTGACACGACACTAT 3457
Db 1943 -----GluAlaAsnGluGlyLeuSerArgGluValSerThrLe 1955
QY 3458 AACTATAACC-----CCGGTCACAACGTCATCCACACGAGGAACCCAA-- 3500
Db 1955 uLysAsnArgLeuArgArgGlyGlyProIleThrPheSerSerArgSerGlyArgAr 1975
QY 3501 -----TCAGTGTCCAGG 3511
Db 1975 gGlnLeuHisIleGluGlyAlaSerLeuGluLeuSerAspAspAlaGluSerLysG1 1995
QY 3512 ACAAGATGGTTCATCTCAGCGGCCTACCC 3542
Db 1995 ySerAspValAsnGluAlaGlnProThrPro 2005
RESULT 5
A61231
myosin heavy chain nonmuscle form A - human
N:Alternate names: cellular myosin heavy chain; myosin type 9; NMVHC-A
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Homo sapiens (man)
C:Date: 12-May-1994 #sequence revision 14-Jul-1994 #text_change 19-Apr-2002
C:Accession: A61231; A34876; I52562; I61692
R:Simons, M.; Wang, M.; McBride, O.W.; Kawamoto, S.; Yamakawa, K.; Gdula, D.; Adelstein,
Circ. Res. 69, 530-539, 1991
A:Title: Human nonmuscle myosin heavy chains are encoded by two genes located on different
A:Reference number: A61231; MUID:91316803; PMID:1860190
A:Accession: A61231
A:Molecule type: mRNA
A:Residues: 1-715 <SIM>
A:Cross-references: GB:M69180; NID:G189029; PIDN:AAA61765.1; PID:G189030
R:Saez, C.G.; Myers, J.C.; Shows, T.B.; Leinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 87, 1164-1168, 1990
A:Title: Human nonmuscle myosin heavy chain mRNA: generation of diversity through altern
A:Reference number: A34876; MUID:90138958; PMID:1967836
A:Accession: A34876
A:Molecule type: mRNA
A:Residues: 715-1961 <SAE>
A:Cross-references: GB:M31013; NID:G189035; PIDN:AAA36349.1; PID:G189036
R:Toothaker, L.E.; Gonzalez, D.A.; Tung, N.; Lemons, R.S.; Le Beau, M.M.; Arnaout, M.A.;
Blood 78, 1826-1833, 1991
A:Title: Cellular myosin heavy chain in human leukocytes: isolation of 5' cDNA clones, c
A:Reference number: I52562; MUID:92003925; PMID:1912569
A:Accession: I52562
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-52, 'EAI', 56-659, 'T', 661-868, 'T', 870-930, 'C', 932-1239, 'KG', 1242-1337 <RES>
A:Cross-references: GB:M81105; NID:G188988; PIDN:AAA59888.1; PID:G553596
R:Bement, W.M.; Hasson, T.; Wirth, J.A.; Cheney, R.E.; Mooseker, M.S.
Proc. Natl. Acad. Sci. U.S.A. 91, 6549-6553, 1994
A:Title: Identification and overlapping expression of multiple unconventional myosin gen
A:Reference number: A55758; MUID:94294418; PMID:8022818
A:Accession: I61692
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 182-218 <BEM>
A:Cross-references: GB:L29141; NID:G457249; PIDN:AAA20904.1; PID:G531134
C:Genetics:
A:Gene: GDB:MYH9
A:Cross-references: GDB:120216; OMIM:160775
A:Map position: 22q12.3-22q13.1
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotid
F:84-764/Domain: myosin motor domain homology <MMOT>
F:174-181/Region: nucleotide-binding motif A (P-loop)
F:552-565/Region: actin binding #status predicted
F:626-640/Region: actin binding #status predicted
F:837-1938/Domain: coiled coil #status predicted <COI>
F:837-1277/Domain: S2 #status predicted <DS2>
F:1278-1961/Domain: light meromyosin #status predicted <LMM>
F:1939-1961/Domain: carboxyl-terminal <CBT>

F:125/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F:180/Binding site: ATP (Lys) #status predicted
F:694,704/Active site: Cys #status predicted
Alignment Scores:
Pred. No.: 1.76e-14 Length: 1961
Score: 469.50 Matches: 228
Percent Similarity: 41.02% Conservative: 181
Best Local Similarity: 22.87% Mismatches: 341
Query Match: 6.07% Indels: 247
DB: 1 Gaps: 40
US-10-788-793-1 (1-4364) x A61231 (1-1961)
QY 273 AAGACTAAGAAGTCTGTGGAGTTATCCAAAGGAGGACCTCATCCAGCTCTGAGTATCATG 332
Db 833 LysValLysProLeuLeuGlnValSerArgGln----- 843
QY 333 GAAGGGAGTTGCGAGGTCGAGAAGATGTCTCATCCACATGCTGAGGACAGAGAAAACCAAG 392
Db 844 GluGluGluMetMetAlaLysGluGluLeu-----ValLysValArg 858
QY 393 CCCGAGGTTCTGGAGGCACACTATGGATCTGCAGAACCTGAGAAAAGTCTTCGGGTCCTG 452
Db 859 GluLysGlnLeuAlaAlaGluAsnArgLeuMetGluMetGluThrLeuGlnSerGlnLeu 878
QY 453 CACCGAGATGCCATCCTTGCTCAAGAGAAAGTCCATAGGAGAACGACGCTCTATGAGAAACCT 512
Db 879 MetAlaGluLysLeuGlnLeuGlnGlnLeuGlnAlaGluThrGluLeuCysAlaGlu 898
QY 513 ATCTCAGAGCTG---GACAGACTGGAGGAAAAGCAGAAGGAGACGTACCGCCGCTGCTGAG 569
Db 899 AlaGluGluLeuArgAlaArgLeuThrAlaLysLysGlnGlu----- 912
QY 570 GAGCAGCTGCTGCTGGCTGAGAAAGTCTCAC--AGCGCACCGGTGTACGAGCTGGAGAAC 626
Db 913 -----LeuGluGluIleCysHisAspLeuGluAlaArgValGluGluGluGlu 928
QY 627 GAGAAGCACAAAGCACACTGACTACATGAACAAGAGCGACGACTTCACCAACCTGCTGGAG 686
Db 929 GluArgTyrGlnHisLeuGlnAlaGluLysLys-----MetGln 942
QY 687 CAGGAGCGGAGAGAGGTTGAAAAAGCTCTTGAACAAGAGAAAAGCTTACCAAGCCCGCAAA 746
Db 943 GlnAsnIleGlnGluLeuGluGluGlnLeuGluGluGluSerAlaArgGlnLysLeu 962
QY 747 GAAAAGGAAAACGCT-----AAGCGGCTCAACAAACTTCGAGATGAGCTTGTGAAG 797
Db 963 GlnLeuGluLysValThrThrGluAlaLysLysLysLysLys----- 982
QY 798 CTCAAGTCTCTTCGCCCTCATGTTGGTGACGAGAGCGCAGATGCACATCGAGCAACTGGGC 857
Db 983 LeuGluAspGlnAsnCysLysLysLeuAlaLysGluLysLys----- 995
QY 858 CTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAAGCTGAGGGAGGAGGAAGAAACTC 917
Db 996 LeuLeuGluAspArgIleAlaGluPheThrThrAsnLeuThrGluGluGluLysSer 1015
QY 918 AAAGCGGTCACTTACAAATCCAAAGGAAGACCGCCAGAAAGCTGCTCAAGTTAGAAAGTGAC 977
Db 1016 LysSerLeuAla---LysLeuLysAsnLysHisGluAlaMetIle-----ThrAsp 1031
QY 978 TTCGAACACAAGGCCTCGAGGTTTCCCAGGAGCACGAAAGAGATGAACCGCAATGGCG 1037
Db 1032 LeuGluGluArgLeuArgArgGluGluLysGlnArgGlnGluLeu----- 1046
QY 1038 AATCAAGAATCTCAAAACCGCAACTTCGACTCAAAACTGGTTGGCTTATCGCAAAGGATT 1097
Db 1047 -----GluLysThrArgArgLysLeuGluGlyAspSerThrAspLeuSerAspGlnIle 1064
QY 1098 GAGGAGCTGGAA-----GAGACCAATAAAAGCCTTCAGAAGGCAGAGGAAGAG 1145
Db 1065 AlaGluLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1084

QY	1146	CTCCAGGAGCTGAGAGAGAAATTGCCAAAGGGGAATGTGGAAACTCCAGTCTCATGGCG	1201
Db	1085	LeuGlnAlaAlaLeuAlaArgValGluGluAlaAlaGlnLysAsnMetAlaLeuLys	1104
QY	1206	GAAGTGGAGAGTCTGCGCAAGCGGTGCTTGAGATGGAG-	1244
Db	1105	LysIleArgGluLeuGluSerGlnIleSerGluLeuGlnGluAspLeuGluSerGluArg	1124
QY	1245	-----GGCAAGGATGAAGAGATCACG	1265
Db	1125	AlaSerArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluLeuGluAlaLeu	1144
QY	1266	AAGACCGAG-----GCCCAGTCCGGGAGCTGAAG-	1295
Db	1145	LysThrGluLeuGluAspThrLeuAspSerThrAlaAlaGlnGlnGluLeuArgSerLys	1164
QY	1296	-----AAGAAGCTCCACAGAGGAACAACACACCAAGCAAGAA	1331
Db	1165	ArgGluGlnGluValAsnIleLeuLysLysThrLeuGluGluAlaLysThrHisGlu	1184
QY	1332	CTTAGACTAGAAGTGAGAAAGCTGCAGAAGAGGATGTCTGAGCTGGAGAAGCTGGAGGAA	1391
Db	1185	AlaGlnIleGln--GluMetArgGlnLysHisSerGlnAlaValGluGluLeuAlaGlu	1203
QY	1392	CGGCTTCAGCCGGAGTAAGTCGGAANTGCACCCAGCTCCATCTGAACCTGGAGAGGAGAAG	1451
Db	1204	GlnLeuGluGlnThrLys-----ArgValLysAlaAsnLeuGluLysAlaLys	1219
QY	1452	-----AACCTAACCAAGACCTGCTGACAGCTGGAGGTG-----	1487
Db	1220	GlnThrLeuGluAsnGluArgGlyGluLeuAlaAsnGluValLysValLeuLeuGlnGly	1239
QY	1488	-----GTCAAGAGTCGAGTTAAAGAACTCGAATGC	1517
Db	1240	GlyArgAspSerGluHisLysArgLysLysValGluAlaGlnLeuGlnGluLeuGlnVal	1259
QY	1518	TCCGAGAGTAGACTGGAGAAGCCGAGTTAAGCCTCAAAAGATGACCTTACAAAGCTGAAG	1577
Db	1260	LysPheAsnGluGlyGluArgValArgThrGluLeuAlaAspLysValThrLysLeuGln	1279
QY	1578	TCCTTCACTGTGATGCTGGTGGATGAGAGGAAAATATATGATGGAGAAAATAAAGCAAGAA	1637
Db	1280	Val-----GluLeuAspAsnValThrGlyLeuLeuSerGlnSer	1292
QY	1638	GAGAGGAAAGTGGATGGTGTGAATAAAACCTTTAAGGTGGAGCAGGGAAAAGTCTATGGAT	1697
Db	1293	AspSerLysSerSerLysLeuThrLysAspPheSerAlaLeuGluSerGlnLeuGlnAsp	1312
QY	1698	GTGACGGAAAAGCTAATCGAGGAAAGCAAGAGCTTTAAAACTCAAATCTGAAATGGAG	1757
Db	1313	ThrGlnGluLeuLeuGlnGluGluAsnArgGlnLysLeuSerLeuSerThrLysLeuLys	1332
QY	1758	GAAGAAGGAGTACAGTCTGACAAAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGAA	1817
Db	1333	GlnValGlu-----AspGluLysAsnSerPheArgGluGlnLeuGluGluGlu	1348
QY	1818	GAAGAAAGGTCC-----TGTAAGCTGAGCTGCAGTGTAGAC	1853
Db	1349	GluGluGluAlaLysHisAsnLeuGluLysGlnIleAlaThrLeuHisAlaGlnValAla	1368
QY	1854	TTACTAAAGAAAGCGG-----CTTGATGGCATAGAGGAGGTAGAA	1892
Db	1369	AspMetLysLysLysMetGluAspSerValGlyCysLeuGluThrAlaGluGluValLys	1388
QY	1893	AGGGAATAAACCGA-----GGTAGGTCGTGCAAGGGGTCTGAGTTACCTGCCCG	1943
Db	1389	ArgLysLeuGlnLysAspLeuGluGlyLeuSerGlnArgHisGluGluLysValAlaAla	1408
QY	1944	GAAGACAATAAGATCAGAGAACTAACGCTTGAATTCGAGAGACTGAAGAAACGGCTCCAG	2003
Db	1409	TyrAsp-----LysLeuGluLysThrLysThrArgLeuGln	1420

QY	2004	CAG---TTGGAG-----GTGGTGGAGGGGAC-----TTG	2030
Db	1421	GlnGluLeuAspAspLeuLeuValAspLeuAspHisGlnArgGlnSerAlaCysAsnLeu	1440
QY	2031	ATGAAGACCGAGGACGAATATGACCAGTTG-----GAGCAGAAG	2069
Db	1441	GluLysLysGlnLysLysPheAspGlnLeuLeuAlaGluGluLysThrIleSerAlaLys	1460
QY	2070	TTCAGAACCGAGCAGGATAAGGCAAACTTCTCTCCAGCAGCTCGAGGAAATCAAACAC	2129
Db	1461	TyrAlaGluGluArgAspArgAlaGluAlaGluAlaArgGluLysGluThrLysAlaLeu	1480
QY	2130	CAAATGGCCCAAGCACAAAGCCATAGAGAAAGGGAGGCCGTGAGCCAGGAAGCCGAACGTG	2189
Db	1481	SerLeuAla-----ArgAlaLeuGlu-----GluAlaMetGluGlnLysAlaGluLeu	1496
QY	2190	CGACAC-----AGTTTCGGCTGGAG-----	2210
Db	1497	GluArgLeuAsnLysGlnPheArgThrGluMetGluAspLeuMetSerSerLysAspAsp	1516
QY	2211	-----GAGGCTAAAGTCTGTGATTACAGGCCGAGGTGCAG	2246
Db	1517	ValGlyLysSerValHisGluLeuGluLysSerLysArgAlaLeuGluGlnValGlu	1536
QY	2247	GCTCTCAAGGAGAAGATCCACGAGCTGATGAACAAG-----GAAGACCAGCTG	2294
Db	1537	GluMetLysThrGlnLeuGluGluLeuGluAspGluLeuGlnAlaThrGluAspAlaLys	1556
QY	2295	TCTCAGCTCCAAGTCGACTATTTCGGTC-----CTTCAGCAA	2330
Db	1557	LeuArgLeuGluValAsnLeuGlnAlaMetLysAlaGlnPheGluArgAspLeuGlnGly	1576
QY	2331	AGATTATTGGAAGAAGAACTAAGAACAAAGACATGGGGAGGGTCTCAATCTGACC	2390
Db	1577	ArgAspGluGlnSerGluGluLysLysGlnLeuValArgGlnValArgGluMetGlu	1596
QY	2391	AAGGAGCTAGACCTTTCCAAGCGCTACGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGA	2450
Db	1597	AlaGluLeuGlu---AspGluArgLysGlnArgSerMetAlaValAlaAlaArgLysLys	1615
QY	2451	AGGATGGTGACGTGCCTGTGGCCTCCACTGGGGTGCAGACCGAGCGGTGTGCGGGAT	2510
Db	1616	LeuGluMetAspLeuLys-----AspLeuGluAlaHisIleAspSer	1629
QY	2511	GCTCGGAGGAGGAGACCCCGGTGTGTTCATTTCGAAATCCTTCAGGAGGAA-----	2564
Db	1630	AlaAsnLysAsnArgAspGluAlaIleLysGlnLeuArgLysLeuGlnAlaGlnMetLys	1649
QY	2565	-----AATCACATCATGAGTAATCTT	2585
Db	1650	AspCysMetArgGluLeuAspAspThrArgAlaSerArgGluGluIleLeuAlaGlnAla	1669
QY	2586	CGACAGGTAGGCCTGAAG--AAACCCATGGAACGGTCTCGGTCTCTCGACAGGTATCCC	2642
Db	1670	LysGluAsnGluLysLysLeuLysSerMetGluAlaGluMetIleGlnLeuGlnGluGlu	1689
QY	2643	CCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTGGATTCTCTTGGATGAGAAAAGAGAA	2702
Db	1690	LeuAlaAlaAlaGluArgAlaLysArgGlnAla-----	1700
QY	2703	AACGGTCTCTCCACTCCGAGGAGAAAGGGCCCCAGGCCAAACACGGGTGCAGGGCACCCC	2762
Db	1701	-----GlnGlnGluArgAspGluLeuAlaAspGluIleAlaAsnSerSer	1715
QY	2763	GGGGAGCTGTCCTAGCACCAAGACGAGGCCAGCCCTACACATCCGTGTG	2813
Db	1716	GlyLysGlyAlaLeuAlaLeuGluGluLysArgArgLeuGluAlaArgIle	1732

RESULT 6

A45592

HA3552
liver stage antigen LSA-1 - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:\pccenter\1\tracmodam\lucaparam
C:\Date: 22-Nov-1993 #sequence revision 02-Dec-1994 #text change 09-Jun-2000

Query Match: 6.01% Indels: 199
DB: 2 Gaps: 32

US-10-788-793-1 (1-4364) x S67593 (1-1790)

```

QY 174 GAAGATGCAAAAAGAACAAAGGCCAATCGGAAGGAG-----GAGGATGTC 218
    ||| ||| ||| : : : : : ||| ||| ||| ||| : : :
Db 889 GluLysAspLysSerAsnValAsnHisGlnLysGluThrLysSerLeuLysGluAspIle 908

QY 219 ATGGCTTCGGAACTATCAAAAGGCACCTCAAAACCATCTCGAGAAAAGTGAAGAAAGACT 278
    ||| ||| : : : : : : : : : : : ||| ||| |||
Db 909 AlaAlaLysIleThrGluIleLysAlaIleAsnGluAsnLeuGluMetLysIleGln 928

QY 279 AAGAAGTCTGTGGAGTTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGG 338
    : : : : : ||| ||| : : : : : : : : : ||| : : :
Db 929 CysAsnAsnLeuSerLysGluLysGluHisIleSerLysGluLeuValGluTyrLysSer 948

QY 339 GAGTTGCAGGCTCGAGAAGATGTATCCACATGCTGAGGACAGAGAAAACCAAGCCCGAG 398
    ||| ||| : : : : : : : : : ||| ||| ||| |||
Db 949 ArgPheGlnSerHisAspAsnLeuValAlaLysLeu--ThrGluLysLeuLysSerLeu 967

QY 399 GTTCTGGAGGCACACTATGGATCTGCAGAACTTGAGAAAGTGTCTCGGTCTCTGCACCGA 458
    ||| ||| ||| ||| ||| : : : : : : : : : : :
Db 968 AlaAsnAsnTyrLysAspMetGlnAlaGluAsnGluSerLeuIleLysAlaValGluGlu 987

QY 459 GATGCCATCCTTGCTCAAGAGAAGTCCATAGGAGAAGACCTCTATGAGAAAACCTATCTCA 518
    ||| ||| ||| ||| ||| : : : : : : : : : |||
Db 988 Ser-----LysAsnGluSerSerIleGlnLeuSerAsnLeuGlnAsnLysIleAsp 1004

QY 519 GAGCTGGACAGACTGGAGGAAAAG-----CAGAAGGAGACGTACCGCCGCATGCTA 569
    : : : : : ||| ||| : : : : : : : : : : :
Db 1005 SerMetSerGlnGluLysGluAsnPheGlnIleGluArgGlySerIleGluLysAsnIle 1024

QY 570 GAGCAGCTGCTGGCTGAGAAGTGTACAGGCGCACCTGTACGAGTGTGAGTGGAGAACGAG 629
    ||| ||| ||| : : : : : : : : : : : ||| |||
Db 1025 GluGlnLeu-----LysLysThrIleSerAspLeuGluGlnThr 1037

QY 630 AAGCACAGCACACTGACTACATGAACAAGCGCAGCTTCACCAACTGCTGGAGCAG 689
    ||| : : : : : ||| ||| ||| ||| ||| : : :
Db 1038 LysGlu-----GluIleIleSerLysSerAspSerSerLysAspGluTyrGluSer 1054

QY 690 GAGCGAGAGAGGTTGAAAAGCTCCTTGAAACAAGAAAAGCTTACCAGCCGCCAAAGAA 749
    : : : ||| ||| : : : ||| ||| ||| ||| |||
Db 1055 GlnIleSerLeuLeuLysGluLysLeu-----GluThrAlaThrThrAlaAsnAspGlu 1072

QY 750 AAGGAAAAC---GCTAAGCGGCTCAACAACTTCGAGATGAGCTTGTGAAGCTCAAGTCC 806
    ||| ||| ||| ||| ||| ||| : : : ||| ||| : : :
Db 1073 AsnValAsnLysIleSerGluLeuThrLysThrArgGluGluLeuGluAlaGluLeuAla 1092

QY 807 TTCGCCCTCATGTTGGTGGACGAGGCGCATGCATCATCGACCACTGGCCCTGCAGAGT 866
    ||| : : : ||| : : : ||| : : : |||
Db 1093 AlaTyrLysAsnLeuLysAsnGluLeuGluThrLysLeuGluThrSerGluLysAlaLeu 1112

QY 867 CAGAAAGTCCAGGACCTCACTCAGAAGCTGAGGGAGGAGGAAGAAAACCTCAAAGCGGTC 926
    : : : : : ||| : : : : : ||| : : : ||| : : : ||| : : :
Db 1113 LysGluValLysGluAsnGluGluHisLeuLysGluGluLysIleGlnLeuGluLys-- 1131

QY 927 ACTTACAAATCCAAGGAGACCGCCAGAGCTGCTCAAGTTAGAAGTGAAGTTCGAACAC 986
    : : : : : ||| : : : ||| ||| ||| ||| : : :
Db 1132 -----GluAlaThrGluThrLysGlnGlnLeuAsnSerLeuArgAlaAsnLeuGlu-- 1148

QY 987 AAGGCCTCGAGGTTTCCAGGAGCACGGAAGAGATGAACGCCAAATTGGCGAATCAAGAA 1046
    : : : ||| ||| ||| ||| : : : : : ||| ||| |||
Db 1149 -----SerLeuGluLysGluHisGluAspLeuAlaAlaGlnLeuLysLysTyrGlu 1165

QY 1047 -----TCTCAACAACCGCAACTTCGACTCAAACTGGTTGGCTTATCGCAAAGG 1094
    : : : ||| ||| ||| ||| : : : : : ||| : : :
Db 1166 GluGlnIleAlaAsnLysGluArgGlnTyrAsnGluGluIleSerGlnLeuAsnAspGlu 1185

QY 1095 ATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAGGCAGAGGAGACTCCAG-- 1151
    ||| : : : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 1186 IleThrSerThrGlnGlnGluAsnGluSerIleLysLysLysAsnAspGluLeuGluGly 1205

```

Qy	1151	-----	-----	-----	1151
Db	1206	GluValLysAlaMetLysSerThrSerGluGluGlnSerAsnLeuLysLysSerGluIle	1225		
Qy	1152	-----	-----	-----	1152
Db	1226	AspAlaLeuAsnLeuGlnIleLysGluLeuLysLysLysAsnGluThrAsnGluAlaSer	1245		
Qy	1197	CTCATGGCGGAAGTGAGAGTCTGCGCAAGCGCGTCTGAGATGGAGGCGAAGGATGAA	1256		
Db	1246	LeuLeuGluSerIleLysSerValGluSerGluThrValLysIleLysGluLeuGlnAsp	1265		
Qy	1257	GAGATCACGAACCGAGGCCAGTGCCGGGAGCTGAAGAAGAAGCTCCAA---GAGGAA	1313		
Db	1266	GluCysAsnPheLysGluLysGluValSerGluLeuGluAspLysLeuLysAlaSerGlu	1285		
Qy	1314	GAACACCACAGCAAG-----	1328		
Db	1286	AspLysAsnSerLysTyrLeuGluLeuGlnLysGluSerGluLysIleLysGluGluLeu	1305		
Qy	1329	-----	-----	-----	1355
Db	1306	AspAlaLysThrThrGluLeuLysIleGlnLeuGluLysIleThrAsnLeuSerLysAla	1325		
Qy	1356	---CAGAAGAGGATGTCTGAGCTGGAGAAGCTG-----	1385		
Db	1326	LysGluLysSerGluSerGluLeuSerArgLeuLysLysThrSerSerGluGluArgLys	1345		
Qy	1386	-----	-----	-----	1439
Db	1346	AsnAlaGluGlnLeuGluLysLeuLysAsnGluIleGlnIleLysAsnGlnAlaPhe	1365		
Qy	1440	GAGAAGGAGAGAACCTAACCAAGAC-----	1466		
Db	1366	GluLysGluArgLysLeuLeuAsnGluGlySerSerThrIleThrGlnGluTyrSerGlu	1385		
Qy	1467	-----	-----	-----	1493
Db	1386	LysIleAsnThrLeuGluAspGluLeuIleArgLeuGlnAsnGluAsnGluLeu-----	1403		
Qy	1494	AGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAGGCCGAGTTAAGCCTC	1553		
Db	1404	---LysAlaLysGluIleAspAsnThrArgSerGluLeuGluLysValSerLeuSer---	1421		
Qy	1554	AAAGATGACCTTACAAAGCTGAAGTCCTTCACCTGTG--ATGCTGGTGGATGAGAGGAAA	1610		
Db	1422	AsnAspGluLeuLeuGluGluLysGlnAsnThrIleLysSerLeuGlnAspGluIleLeu	1441		
Qy	1611	AAATATGATGGAGAAAAATAAAGCAAGAGAGAGAGAAAGTGGATGGTTGAATAAAAACTTT	1670		
Db	1442	SerTyrLysAspLysIleThrArgAsnAspGluLysLeuLeuSerIleGluArgAspAsn	1461		
Qy	1671	AAG-----	1712		
Db	1462	LysArgAspLeuGluSerLeuLysGluGlnLeuArgAlaalaGlnGluSerLysAlaLys	1481		
Qy	1713	ATCGAGGAACCAAGACTTTTAAAACTCAAATCTGAAATGGAGGAAAAGGAGTACAGT	1772		
Db	1482	ValGluGluGlyLeuLysLysLeuGluGluGluSerLysGluLysAlaGluLeuGlu	1501		
Qy	1773	CTGACAAAGGAGGAGGATGAGCTGATGGGTAAACTGAGGAGCGAA---GAAGAAAAGGTCC	1829		
Db	1502	LysSerLys-----GluMetMetLysLysLeuGluSerThrIleGluSerAsnGlu	1518		
Qy	1830	TGTGAAGTACGAGTGTAGACTTACTAAAGAACGGCTTGATGGCATAGAGGAGGTA	1889		
Db	1519	ThrGluLeuLysSerSerMetGluThrIleArgLysSerAspGluLysLeuGluGln---	1537		
Qy	1890	GAAAGGGAATAAACCGAGGTAGTCTGTGCAAGGGGTCTGTAGTTACCTGCCCGGAAGAC	1949		
Db	1538	-----	-----	-----	1545
Qy	1950	-----	-----	-----	1973

Db 1546 IleLysAsnLeuGlnHisGluLysSerAspLeuIleSerArgIleAsnGluSerGluLys 1565
QY 1974 GAAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGTGGTGGAGGGGACTTGATG 2033
Db 1566 AspIleGluGluLysSerLysLeu--ArgIleGluAlaLysSerGlySerGluLeu 1584
QY 2034 AAGACC---GAGGACGAATATGACCAGTTGGAGCAGAAGTTCAGAAACCGAGCAGGATAAG 2090
Db 1585 GluThrValLysGlnGluLeuAsnAsnAlaGlnGluLysIleArgIleAsnAlaGluGlu 1604
QY 2091 GCAAACTTCCTCTCCAGCAGCTCGAGGAATCAAACACCAATGGCCAAGCACAAAGCC 2150
Db 1605 AsnThrValLysSerLysLeuGluAspIleGluArgGluLeuLysAspLysGlnAla 1624
QY 2151 ATAGAGAAAGGGAGCGCTGAGCCAGGAAGCCGAACTGCGACACAGGTTTCGGCTGGAG 2210
Db 1625 GluIleLysSerAsnGlnGluLysGlu-----LeuLeu 1636
QY 2211 GAGGCTAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAGGAGAAGATCCACGAG 2270
Db 1637 ThrSerArgLeuLysGluLeuGluGlnGluLeuAspSerThrGlnGlnLysAlaGlnLys 1656
QY 2271 CTGATGAACAAGGAAGAC-----CAGCTGTCTCAGCTCCAAGTCGACTATTTCGGTC 2321
Db 1657 -----SerGluGluGluArgAlaGluValArgLysPheGlnValGluLysSerGln 1674
QY 2322 CTTCAGCAAAAGATTTATGGAAGAAAGAACTAAG-----AACAAAGAACATG 2366
Db 1675 LeuAspGluLysAlaMetLeuLeuGluThrLysTyrAsnAspLeuValAsnLysGluGln 1694
QY 2367 GGG-----AGGAGGTC 2378
Db 1695 AlaTrpLysArgAspGluAspThrValLysLysThrAspSerGlnArgGlnGluIle 1714
QY 2379 CTTCAATCTGACCAAGGAGCTAGAGCTTTCCAAAGCGCTACAGCCGAGCTCTCAGGCCGAGT 2438
Db 1715 GluLysLeuAlaLysGluLeuAspAsnLeuLysAlaGluAsnSerLysLysLeuLysGluAla 1734
QY 2439 GGGAAACGGCCGAGGATGTTGGAC 2462
Db 1735 AsnGluAspArgSerGluIleAsp 1742

RESULT 8
T21174
hypothetical protein F20G4.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C;Accession: T21174; T25113
R;McMurray, A.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19385
A;Accession: T21174
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2020 <WIL>
A;Cross-references: EMBL:Z75538; PIDN:CAA99841.1; GSPDB:GN00019; CESP:F20G4.3
A;Experimental source: clone F20G4
R;McMurray, A.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19982
A;Accession: T25113
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2020 <W12>
A;Cross-references: EMBL:Z75550; PIDN:CAA99931.1; GSPDB:GN00019; CESP:F20G4.3
A;Experimental source: clone T22C1
C;Genetics:
A;Gene: CESP:F20G4.3
A;Map position: 1
C;Introns: 133/3; 155/2; 195/2; 270/3; 325/1; 1023/3; 1072/2; 1125/2; 1207/3; 1248/3; 13
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;106-798/Domain: myosin motor domain homology <MMO>

Alignment Scores: 3.34e-14 Length: 2020
Pred. No.: 463.50 Matches: 268
Score: 38.69% Conservativity: 233
Best Local Similarity: 20.69% Mismatches: 443
Query Match: 6.00% Indels: 351
DB: 2 Gaps: 52
US-10-788-793-1 (1-4364) x T21174 (1-2020)
QY 237 AAAAGGCACCTCAAACCATCTGGAGAAAAGTGAGAAAAAGACT----- 278
Db 885 GluGlnGluLeuLysValThrAlaGluLysLeuArgSerGluValPheIleSerAsp 904
QY 279 ---AAGAAAGTCTGTGGAGTTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAA 335
Db 905 TyrLysGlnGlnMetGluLysMetAspGluGluArgLeuValLeuLysThrArgLeuAsp 924
QY 336 GGGGAGTTGCAGGCTCGAGAAGATGTCTATCCACATCTGAGGACAGAGAAAACCAAGCCCC 395
Db 925 AlaGluSerSerGluArgAlaGluIlePheGluGluArgSerArgMetAlaAlaArgArg 944
QY 396 GAGGTTCTGGAGGCACACTATGGATCTGCAGAACCTTGAGAAAAGTGTTCGGGCTCTGCAC 455
Db 945 AspGluLeuGluGly-----IleLeuGluGluValSer 955
QY 456 CGAGATGCCATCCTTGCTCAAGAGAAGTCCATAGGAGAAGACGTCTATGAGAAAACCTATC 515
Db 956 LysArgLeuGluIleGluGlnLysAlaLysLysAlaAspSerGluSerArgLysLeu 975
QY 516 TCAGAGCTGGACAGACTGGAGGAAGACAGAGGAGAGACGTACCCGCCATGCTAGAGACAG 575
Db 976 ThrGluMetValArgHisLeuGluGluAsnLeuGluAspGluGluArgSerArgGlnLys 995
QY 576 CTGCTGCTGGCTGAGAAGTGTACACAGCGCCACCGTGTACGAGCTGGAGAACGAGAGACAC 635
Db 996 LeuLeuLeuGluLysAsnSerIleGluSerArgLeuLysGluLeuAlaGlnGlyLeu 1015
QY 636 AAGCACACTGACTACATGAACAAG-----AGC 662
Db 1016 GluLeuGluAspSerGlyAsnLysLeuSerLysLysAlaLeuGluArgCys 1035
QY 663 GACGACTTCACCAACCTGCTGGAGCAGGAGCAGAGAGGTTGAAAAAGCTCCTT----- 716
Db 1036 GluAspLeuSerSerArgLeuIleAspGluValGluArgSerLysGlnLeuValLysAla 1055
QY 717 -----GAACAAGAAAAAGCTTACCAA 737
Db 1056 LysAlaArgLeuGluAlaThrValAlaGluIleAsnAspGluLeuLysGluLysGln 1075
QY 738 GCCCGCAAAGAAAGAAAACCGTAAGCGGCTC-----AACAACTTCGAGATGAG--- 788
Db 1076 GlnArgHisAsnAlaGluThrAlaArgArgAlaAlaGluThrGlnLeuArgGluGln 1095
QY 789 -----CTTGTGAAGCTCAAGTCCTTCGCCCTCATGTTGGTGGACGAG----- 830
Db 1096 GluSerCysLeuGluLysThrArgLys---AlaGluGluLeuThrAsnGlnLeuMetArg 1114
QY 831 AGGCAGATGCACATCGAGCAACTGGGCCCTGCAGAGTCAGAAAAGTCCAGGACCTCACTCAG 890
Db 1115 LysGluSerGluLeuSerGlnIleSerIleArgAsnAspGluGluLeuAlaAlaArgGln 1134
QY 891 AAGCTGAGGAGGAGGAAGAAAAACTCAAAGCG-----GTCACT 929
Db 1135 GlnLeuGluArgGluIleArgGluIleArgAlaGlnLeuAspAspAlaIleGluGluThr 1154
QY 930 TACAAATCCAAAGGAAGACCGCCAGAGCTGTCAAGTTAGAGTGGACTTCGAACACAAG 989
Db 1155 AsnLysGluLysAlaAlaArgGlnLysAlaGluLysAlaArgArgAspMetAlaGluGlu 1174
QY 990 GCCTCGAGGTTTCCCGAGGAG-----CACGAAGAG 1019
::: |||||

Db 1175 LeuGluSerTyrLysGlnGluLeuGluSerAsnAspLysThrValLeuHisSerGln 1194
QY 1020 ATGAACGCCAAATTGGCGAATCAAGAATCTCAC---AACCGCAAACTTCGACTCAAAGCTG 1076
Db 1195 LeuLysAlaLysArgAspGluGluTyrAlaHisLeuGlnLysGlnLeuGluGluThrVal 1214
QY 1077 GTTGGCTTATCGCAAGAGATTGAGGAGCTGGAAGAGACCAAT---AAAAGCTTCAGAAG 1133
Db 1215 LysSerSerGluGluValValGluGluMetLysAlaGlnAsnGlnLysLysIleGluGlu 1234
QY 1134 GCAGAGGAAGAGCTCCAGGAGCTG---AGAGAGAAAATT-----GCC 1172
Db 1235 LeuAsnGluThrIleAspGlnLeuLysArgGlnLysIleSerAlaAspLysAlaLysSer 1254
QY 1173 AAAGGGGAATGTGGAACCTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCTG 1232
Db 1255 SerAlaGluSerAspAsnGluAsnPheArgAlaGluLeuSerAsnIleAlaSerAlaArg 1274
QY 1233 CTTGAGATGGAGGGCAAGGATGAAGAGATCAGGAAGACCGAGGCCAGTCCGGGAGCTG 1292
Db 1275 LeuGluAlaGluLysLysArg-----LysAlaAla 1284
QY 1293 AAGAAGAAAGCTCCAAGAGGAAGAACACACAGCAAGGAACCTTAGACTAGAGTGGAGAAG 1352
Db 1285 GluThrSerLeuMetGluLysAspHisLysMetArgGluMetGlnSerAsnLeuAspAsp 1304
QY 1353 CTGCAGAAAGAGGATGTCTGAGCTGGAGAGCTGGAGGACGTTTCAGCCGAGTAAGTCG 1412
Db 1305 LeuMetAlaLysLeuSerLysMetAsnAsnGluLeuGluSerIleGlnLysAlaLysSer 1324
QY 1413 GAATGCACCCAGCTCCATCTGAACCTGGAGAAGGAGAGAACCTTAACCAAGACCTG--- 1469
Db 1325 AlaAspGluThrLeuAsnSerAsnLeuLys---LysAsnAlaSerLeuAspMetGln 1343
QY 1470 CTGAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGTCTCCGAGAGTAGA 1529
Db 1344 LeuSerGluLeuThrGluAlaSer-----GluGluAspArg 1355
QY 1530 CTGGAGAAGCCGAGTTAAGC-----CTCAAAGATGACCTTACAAAGCTG 1574
Db 1356 ArgThrArgAlaThrLeuAsnAsnLysIleArgGlnLeuGluGluAspLeu----- 1372
QY 1575 AAGTCCTTCACCTGTGATGCTGGTGGATGAGAGGGAATAATATGATGGAGAAAAATAAGCAA 1634
Db 1373 ---AlaValAlaValGluAlaArgAspAlaLeuAspAlaGlnGluLysIleGluLys 1391
QY 1635 GAAGAGAGGAAGTGGATGGGTTGAATAAAACCTTTAAGGTGGAGCAGGGAAGTCAATG 1694
Db 1392 GluValLysGluValLysSerLeu-----LeuAla 1401
QY 1695 GATGTGACGGAAGAAAGCTAATCGAGGAAGCAAGAGCTTTTAAAACTCAAATCTGAAATG 1754
Db 1402 GluAlaArgLysLysLeuAspGluGluAsnArgGluValMetGluGluLeuArgLysLys 1421
QY 1755 GAGGAAGAGGAGTACAGTCTGACAAAGGAGAGGATGAGCTGATGGGTAAACTGAGGAGC 1814
Db 1422 LysGluLysGluLeuSerAlaGluLysGluArgAlaAspMetAlaGluGlnAlaArgAsp 1441
QY 1815 GAAGAAGAAAGGTCCTGTGAACCTGAGCTGCAGTGTAGACTTACTAAAGAAGCGGCTTGAT 1874
Db 1442 LysAlaGluArgAla-----LysLysLysAlaIleGln 1452
QY 1875 GGCAATAGAGGAGGTAGAAAGGGGAATAAAACCGAGGTAGGTCTGCAAGGGGTCTGAGTTC 1934
Db 1453 GluAlaGluAspValGlnLysGluLeu----- 1461
QY 1935 ACCTGCCCGGAAGACAAATAAGATCAGAGAACCTAACGCTTGAAATCGAGAGA---CTGAAG 1991
Db 1462 -----ThrAspValValAlaAlaThrArgGluMetGluArgLysMetArg 1476
QY 1992 AAACGCTCCAGCAGTTGGAGGTGGTGGAGGGGACCTTGATGAAGACCGGAGGACCAATAT 2051
Db 1477 LysPheAspGlnGlnLeuAlaGluGluArgAsnAsnThrLeuLeuAlaGlnGlnGluArg 1496

QY 2052 GACCAAGTTGGAGCAGAAAGTTCAGAAACCGAGCAGGATAAAGCAAAATTCCTCTCCAGCAG 2111
Db 1497 AspMetAlaHisGlnMetLeuArgAspAlaGluThrLysAlaLeuValLeuSerAsnGlu 1516
QY 2112 CTCGAGGAA-----ATCAAACACCAAAATGGCCCAAGCACAAAGCCATAGAGAAAGG 2162
Db 1517 LeuSerGluLysLysAspIleValAspGlnLeuGluLysAspLysArgThrLeuLysLeu 1536
QY 2163 GAG-----GCCGTGAGCCAGGAAGCCGAACCTCGCACACAGGTTTCGGCTGGAG 2210
Db 1537 GluIleAspAsnLeuAlaSerThrLysAspAspAlaGlyLysAsnValTyrGluLeuGlu 1556
QY 2211 GAGGCTAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAGGAGAGATCCACGAG 2270
Db 1557 LysThrLysArgArg---LeuAspGluGluLeuSerArgAlaGluGlnIleIleGlu 1575
QY 2271 CTGATGAAC-----AAGGAAGACACGAGCTGTCTCAGCTCCAAGTCGACTATTTCG 2318
Db 1576 LeuGluAspAlaLeuGlnLeuAlaAspAspAlaArgSerArgValGluValAsnMetGln 1595
QY 2319 GTCCTTCAGCAAAAGATT-----ATGGAAGAAGAAACTAAG 2354
Db 1596 AlaMetArgSerGluPheGluArgGlnLeuAlaSerArgGluGluAspGluAspArg 1615
QY 2355 AACAAAGAACATGGGAGGGAGGTCTCAATCTGACCAGGAGCTAGAGCTTTCCAAAGCGC 2414
Db 1616 LysLysGlyLeuThrSerLysIleArgAsnLeuThrGluGluLeuGluSerGluGlnArg 1635
QY 2414 ----- 2414
Db 1636 AlaArgGlnAlaAlaIleAlaAsnLysLysLysIleGluSerGlnIleSerGluLeuThr 1655
QY 2415 -----TACAGCCGAGCTCTCAGGCCGAGT 2438
Db 1656 GluLysAsnGluAlaSerLeuArgGlnIleGluAspLeuSerArgGlnLeuArgLysAla 1675
QY 2439 GGGAAAGGCCGAAGG-----ATGETGGACGTGCCTGTGGCTCCACTGGGGTGCAGACC 2492
Db 1676 GlnLeuGlyTrpLysAspLeuGlnLeuAspValThrGluAlaArgAlaAlaMetGlu-- 1694
QY 2493 GAGGCGGTGTGCGGGGAT-----GTCGCGGAGGAGGAGACC 2528
Db 1695 AspAlaLeuAlaGlyGlnArgAspAlaGluLysArgAlaArgAlaSerGluAspGluIle 1714
QY 2529 CCGGTGTGTTCAAT-----CGCAAATCCTTCCAGGAG 2561
Db 1715 LysArgLeuThrAlaAspIleGlnAlaValSerSerLysArgLysAlaGluAlaGlu 1734
QY 2562 GAAATCACATC-----ATGAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCCATG 2612
Db 1735 ArgAspGluLeuIleGluGluValSerSerLeuArgAlaSerSerPheSerAsnGluGlu 1754
QY 2613 GAACGG-----TCCTCGGTCTCTCGAC-----AGGTATCCCCCAGCA 2648
Db 1755 LysArgArgLeuGluAlaLysValIleAspLeuGluAspGlnLeuAspGluGluAlaSer 1774
QY 2649 GCGAATGAGCTCACCATGAGGAAGTCTTGGATTCTTGGATGAGAAAAAGAGAAACCGGT 2708
Db 1775 AlaAsnGluLeuAlaGlnGluLys-----ValArgLysSerGln----- 1787
QY 2709 CCTTCCACTCCGCAAGGAGAAAGGGCCCCAGGCCAACACAGGGTGCAGGGCACCCCGGGGAG 2768
Db 1787 ----- 1787
QY 2769 CTGGTCTTAGCACCAAAGCAGGGGCCAGCCCCCTACACATCCGTGTGACACCAGATCATGAG 2828
Db 1788 -----GlnGlnLeu-----Glu 1791
QY 2829 AACAGCACTGCCACCTGGAGATCACAAAGCCCCACATCTGAAGAGTTTTCTCTAGTACC 2888
Db 1792 GlnMetThrAlaAspLeuAlaMetGluArgSerValCysGluArg----- 1806

QY 2889 ACCGTCATTCCTACCTTAGGCAACACAGAAACCAAGATAACCAATTATTCCATCACCCAAT 2948
Db 1807 -----ThrGluSerAspLysIleAlaLeuGluArgAlaAsnArg 1819
QY 2949 GTCATGTCGCAAAAGCCCAAAAGTGCAGATCCTACTCTCGGCCCAAGACGACCATGTCC 3008
Db 1820 AspLeuLysGlnGlnLeuGlnAspAlaGluAsn----- 1830
QY 3009 CCTGTCACGATTACTACTATTTCACAGAGAGAGAGCCCGGAAGGTGGAAGGCGCCTTT 3068
Db 1831 -----ThrAlaValAlaArgLeuArgThr----- 1838
QY 3069 GCCGACAGGCCTGCATCCCCCATCCAAATCATGACGGTGTCAACATCTGCAGCTCCCACT 3128
Db 1839 -----GlnIleAsnValAlaGluAlaLysValSerSerLeu 1850
QY 3129 GAAATCGCTGTCTCTCGAATCTCAGGAA---GTGCCTATGGGAAGGACTATC----- 3179
Db 1851 GluGlnGlnLeuSerLeuGluGluGlnAspLysMetArgGlnGlyArgThrLeuArgArg 1870
QY 3180 -----CTCAAAGTCACCCCGGAAACAAACTGTTCCAGCCCCCGTGCGGGAAGTACAAC 3233
Db 1871 MetGluThrLysMetAlaGluMetGlnGlnMetLeuGluGluLysArgGlnGlyGlu 1890
QY 3234 TCCAATGCTAATATCATCACACCGAAGACATAAAATT----- 3272
Db 1891 SerAsnArgGlnAlaValAspArgGlnAsnAlaArgIleArgGlnLeuArgThrGlnLeu 1910
QY 3273 -----CACATTCACTGGGTTCTCAGTTTAAGCGATCTCCTGGGCCT 3314
Db 1911 GluAspThrGluAlaGluArgAspArgLeuThrAsnLysLeuLysAspGluArgArg 1930
QY 3315 GCCGCTGAAGCGGTGAGCCCGATTATCACCGTCCGGCTGTCAACGTGACAGCGGAGAAG 3374
Db 1931 AlaGluGlu-----MetThrAspLeuAsnGluThrLeuSerArg 1943
QY 3375 GAGGTTTCT-----ACAGGCACAGTCTCTCGCTCTCCCAGG-----AAC 3413
Db 1944 AspValSerLeuLeuLysGlnArgGluThrThrAlaArgArgThrProGlyLeuIleGly 1963
QY 3414 CACCTCTCTTCAAGACCCGGTGTAGCAAAAGTGACCAAGCTATAACTATAACCCCGGTC 3473
Db 1964 HisArgGluSerArgArgPheGlySerAsn-----ThrSerLeu 1976
QY 3474 ACAACGTATCCACACGAGGAACCAATCAGTGTGAGGACAAGATGGGTCACTCAGCGG 3533
Db 1977 AlaArgAspGluPheArgGlySer--AlaLeuThrAsnGluMetSerProSerAspArg 1995
QY 3534 CCTACCCCCACCCGCATTCTCTATGTCAAAAGGTATGAAAGCTGGA 3578
Db 1996 ProAla---SerArgLeuThrSerGlyThrGlySerGlnLeuGly 2009

RESULT 9
T24806
hypothetical protein T10G3.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24806
R;Burton, J.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19937
A;Accession: T24806
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1164 <WIL>
A;Cross-references: EMBL:Z81118; PIDN:CAB03330.1; GSPDB:GN00023; CESP:T10G3.5
A;Experimental source: clone T10G3
C;Genetics:
A;Gene: CESP:T10G3.5
A;Map position: 5
A;Introns: 15/3; 73/3; 387/3; 412/3; 708/3; 818/3; 894/3; 982/3; 1080/2
Alignment Scores:

Pred. No.: 3.48e-14 Length: 1164
Score: 462.50 Matches: 241
Percent Similarity: 41.58% Conservative: 196
Best Local Similarity: 22.93% Mismatches: 344
Query Match: 5.98% Indels: 270
DB: 2 Gaps: 50
US-10-788-793-1 (1-4364) x T24806 (1-1164)
QY 30 GCGAAAGAACACACACCATTTGGTTAAGGAGTCGACAACAGGTGGGAATGAGATCACGAAAT 89
Db 156 GlyMetValThrArgGlnPheLeuGluLeuGluLysGluSerGlyGlnGlnSerArgGlu 175
QY 90 CAAGGTGGAGAAAGTTCATCTAAACGGGCATGTCTCTCTGCCCAAGTCCTCCATCATC--- 146
Db 176 LeuGlnGlnValLysGlnGluArgGlyAspLeuMetAlaLysLeuLysGlnMetSerVal 195
QY 147 -----AGCAGTGTATGGTGGTAAGGCCCTCAGAAAGATGCAAAA 185
Db 196 ThrMetArgGluIleThrAspGluSerGluSerGlyLysValGluMetGluAspLeuLys 215
QY 186 AAGAACAGGCCAATCGGAAGGAGGAGGATGTCTGCTTCCGGAACCTATCAAAAGGCAC 245
Db 216 -----ArgGluLeuLysValValLysSerAspValValArgTyrGlu 229
QY 246 CTCAAAACCATCTGGA-----GAAAGTGAGAAAAGACTAAGAAAGTCTGTGGAG 293
Db 230 IleGluValSerArgLeuGluLysMetLeuAspGlnArgProSerGluAspAspValAsn 249
QY 294 TTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGG----- 338
Db 250 ValLeuArgThrGluLeuValAsnAlaGlnLysLeuMetAspAlaIleSerGlnGluLys 269
QY 339 -----GAGTTGCAGGCTCGAGAAGATGTCTATCCACATGCTGAGGACAGAGAAACCAAG 392
Db 270 AspIleGluIleLysGluHisLeuAsnSerIleArgAsnLeuSerMetGluArgGluLys 289
QY 393 CCCGAGGTT-----CTGGAGGCACACTATGGATCTGCAGAACCTGAG---AAA 437
Db 290 GlnHisIleValAsnGluAsnLeuGluLysIleGlyGluGlyGluThrValLys 309
QY 438 GTGCTTCGGGTCCTGCAC-----CGAGATGCCATC 467
Db 310 GlnLeuGlnIleSerTyrAspAlaGlnSerGluGluLeuLysGlnArgAsnGluArgVal 329
QY 468 CTTGCTCAAGAGAAAGTCCATAGGAGAAAGCGTCTATGAGAAACCTATCTCAGAG----- 521
Db 330 ValGlnLeuGluAlaArgIleGluGluAsnValPheGlu-----LeuSerGluAsnLys 347
QY 522 -----CTGGACAGACTGGAGGAAAAGCAGAGGAGACGTAC-----CGCCGCTGCTA 569
Db 348 GlnAsnValLysArgLeuGluAspLysValGlnGluSerGlnAspAlaLeuGlnMetLeu 367
QY 570 -----GAGCAGCTGCTG---CTGGCTGAGAAAGTGTACAGG 602
Db 368 SerAsnIleAsnGlySerAsnGluGluGlnMetIleSerLeuAsnSerLysPheGluArg 387
QY 603 CGCACC-----GTGTACGAG-----CTGGAGAAC 626
Db 388 AsnThrAlaGluArgLysArgIleGluAlaValPheGluGluLysValThrValGlnGly 407
QY 627 GAGAAGCACAAAG-----CACACTGACTACATGAACAGACGACGACTTC 671
Db 408 GluArgLeuLysThrLeuGluMetAlaAsnLeuAspLeuThrAsnGlnLeuAlaSerMet 427
QY 672 ACCAACCTGTGAGCAGGAGCGA-----GAGAGGTTGAAAAGCTCCTTGAA 719
Db 428 GlySerLeuLeuAspLysGluArgSerLeuLeuGluGluLysAsnLysGluIleSerGlu 447
QY 720 CAAGAAAAAGCTTACCAAGCCCGCAAGAAAAG-----GAAAAACGCTAAGCGGCTCAAC 773
Db 448 ArgAspSerSerIleAsnAspLeuLysGluLysLeuAlaGluSerGluLysAlaThr 467

QY 774 AAACCTTCGAGATGAGCTTGTG----- 794
Dbb LysTyrLysAsnGluLeuLysGluHisAlaAspLeuValGluAsnLeuThrLeuGlnLeu 487
QY 795 ---AAGCTCAAGTCTTCGCCCTCATGTTGGTGGACGAG----- 830
Dbb AsnLysLeuGlnGluAsnSerLysAspLeuMetGluLysIleSerAlaGlyGluGlyGly 507
QY 831 AGGCAGATGCACATCGAGCAACTG----- 854
Dbb AlaLysMetAlaIleGluGlnLeuGluGlnLysValLysLeuThrAsnGluLeuGln 527
QY 855 -----GGCCTGCAGAGTCAGAAAGTCCAGGACCTC 884
Dbb ThrSerSerGluLysThrLysLysAlaSerGlyGluLeuAlaLysIleSerGluLeu 547
QY 885 ACTCAGAAAGTGGAGGAGGAAGAAAACTCAAAGCGGTCACTTACAAATCCAAGGAA 944
Dbb LysLysLysLeuArgAspAlaGluAlaSerArgThrAspLysGluGlnLysTrpLysGln 567
QY 945 GACCGCCAG-----AAGCTGCTCAAGTTAGAGTGGACTTCAACACAAAGGCC 992
Dbb GluLysGluSerPheGluArgLysLeuAlaGluAlaGluAspGluIleLysArgLysGly 587
QY 993 TCGAGGTTTCCAGGAGCACGAAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCAC 1052
Dbb GluArgPheValGluMetGluLysGluMetGluGlu-----ArgGlnLysAlaThr 605
QY 1053 AACCGGCAACTTCGACTCAAA-----CTGGTTGGCTTATCGCAAAGGATTGAG----- 1100
Dbb AspArgThrLeuLysLeuLysAspAlaLeuValAsnSerGluLysAsnLeuGluThrIle 625
QY 1101 -----GAGCTGGAAGAGACCAATAAAAGCCCTTCAGAGGCAGAGGAGACTCCAGGAG 1154
Dbb LysLysGluSerGluAspArgGluLysIleValArgGluLysAspAlaHisLeuGluGlu 645
QY 1155 CTGAGAGAGAAATTGCCAAAGGGGAATGTGGAACCTCCAGTCTCATGTGGCGGAAGTGGAG 1214
Dbb AsnLysLysArgIle-----GluAspAlaValGlnLysLeuGlu 658
QY 1215 AGTCTGCCGAAGCGCTGCTTGATGAG-----GGCAAGGATGAAGAGATC 1262
Dbb GluAlaGluLysArgAlaArgGluLeuGluAlaSerValSerSerArgAspThrThrVal 678
QY 1263 ACGAAGACCGAGGCCAGTCCGGGAGCTGAAGAAGACTCCAAGAGGAAGAACACCAC 1322
Dbb SerThrLysGluSerGluLeuSerGluLeuLysGlyLysLeuThrGluSerAsnSerPhe 698
QY 1323 AGCAAGAACTTAGACTAGAAGTGGAGAAGCTGCAGAAGAGGATGTCT----- 1370
Dbb IleGluGluLeuLysValGlnValGluLysValSerAsnGluIleSerGluLysGlnGln 718
QY 1371 GAGCTGGAGAAGCTG----- 1385
Dbb GluValGluAsnLeuMetAlaGluMetArgAspLysGluAlaHisTrpLysThrLysArg 738
QY 1386 -----GAGGAAGCGTTTCAGCCGG 1403
Dbb AspGluPheGluAlaGlnMetLeuArgAsnGlnGluAspAsnGluGluAlaSerSerThr 758
QY 1404 AGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAGGAGAAGAACCTAACCAA 1463
Dbb LeuLysSerValGlnGluGln-----LeuMetLysGluLysGluThrSerGly 774
QY 1464 GACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAG 1523
Dbb GluGluLysAsnGlnLeuIleSerValLysSerGlnLeuGluGluLeuLysThrGluVal 794
QY 1524 AGTAGACTGGAGAGCCGAGTTAAGCCCTCAAAGATGACCTTACAAAGCTGAAGTCCTTC 1583
Dbb GluArgLeuIleArgSerGluGluLysThrGlnGluIleGluLysLeuLysSerAla 814
QY 1584 ACTGTGATGCTGCTGGATGAGAGGAAAAAATATG-----ATGGAGAAAAATAAGCAA 1634

Dbb ValThrAlaThrThrGlnGluArgAspGluLeuThrAlaThrSerGluSerLeuArgThr 834
QY 1635 GAA-----GAGAGGAAAGTGGATGGGTGGAATAAAAAACTTTAAGGTGGAGCAG 1682
Dbb GluCysGluAsnLeuAsnSerLysIleGlnSerIleGluGluSerArgArgHisAlaGlu 854
QY 1683 GGAAAAGTCATGGATGTGACGGAAAAGCTAATCGAGGAAAGCAAGAGCTTTTAAACTC 1742
Dbb GluLysGlySerGluAsnLeuGluArgMetIleThrGluLysSerArgLeuGluLys--- 873
QY 1743 AAATCTGAAATGGAGGAAAAAGGAGTACAGTCTGACAAAGGAGAGGGATGAGCTGATGGGT 1802
Dbb -----AspIleGluGluArgGluSerThrIleGlnSerIleGlnGluAlaLeuGluThr 891
QY 1803 AAACCTGAGGAGCGAAGAAAGGTCTGT-----GAACTGAGC 1841
Dbb LysAspAsnGluIleGluSerLeuLysThrThrGlnArgValValGluAspGluLeuVal 911
QY 1842 TGCAGGTGTAGACTTTACTAAAGAAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGGGAATA 1901
Dbb SerLysIleSerHisIleGluSerPheAsnSerArgIleGluGluPheGluLysGluMet 931
QY 1902 AACCGAGGTAGTCTGTC-----AAGGGTCTGAG----- 1931
Dbb AlaSerGlyLysArgThrIleGluArgLeuGluAlaGluLysAlaGluGluThrGluLys 951
QY 1932 -----TTCACTGCCCGGAAGACAAATAAGATCAGAGAACTAACGCTTGAAATCGAG 1982
Dbb LeuValValPheThrGlyThrGlnSerGlnLysGlnGluGluLeu-----Glu 967
QY 1983 AGACTGAAGAAACCGCTCCAGCAGTTGGAGGTGGTGGAGGGGACTTGATGAAGACCGAG 2042
Dbb LysLeuGlnLysGluIleGlnGluLysGluThrThr-----IleAlaArgMetThr 984
QY 2043 GACGAATATGACCAGTTGGAGCAGAAAGTTTCAGA----ACCGAGCAGGATAAGGCAAACTTC 2099
Dbb SerSerLysThrGlnPheGluAlaMetPheAlaAspValGlnGlnThrLeuSerLysGlu 1004
QY 2100 CTCTCCCGCAGCTCGAGGAAATCAAACACCAATGGCCCAAGCACAAAGCCATAGAGAAA 2159
Dbb IleAsnAspLysThrGluGluIleGluArgLeuMetGluArgIleAspSerLeuGluLys 1024
QY 2160 GGGAGGCCGTGAGCCAGGAAGCGGAACCTGCGACACAGGTTTCGGCTGGAGGAGGCTAAA 2219
Dbb -----ValAsnHisSer---ArgIleGluGluLeuGlu 1034
QY 2220 AGTCGTGATTTACAGGCCGAG-----GTGCAGGCTCTCAAGGAGAAGATCCACGAGCTG 2273
Dbb SerArgLeuThrGlnArgGluArgValValGluSerLeuGluAlaAspLeuAlaVal 1054
QY 2274 ATGAAC--AAGGAAGACCCAGCTGTCTCAGCTCCAAGTCGACTATTTCGGTCCCTTCAGCAA 2330
Dbb ArgAsnIleGluGlnGluLysLeuAspGluLeuGln-----LysLeuLysGlu 1070
QY 2331 AGATTTATGGAAGAAGAAACTAAGAACAAAGACATGGGGAGGAGGTCTCTCAATCTGACC 2390
Dbb GluPhe-----AspGlyLysPheLysTyrPheLysSerArgIleLeuPheAsnSer 1087
QY 2391 AAGGAGCTAGAGCTT--TCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGC 2447
Dbb LysLysLeuGlnIleAspAlaAspArgTyrSer-----Ser 1099
QY 2448 CGAAGGATGTTGGACGTGCTGTGGCCTCCACTGGGTGCAGACCGAGGCGGTG----- 2501
Dbb ArgLysTrpLeuAsp-----AspAlaGluAlaIleAsnCys 1111
QY 2502 -----TGCGGGG-----ATGCTGCGGAGGAGAGACCCCGGCTGTGTTTC 2540
Dbb ThrGluCys-GlyLysValPheSerLeuThrValArgLysHisHisCysArgValCysGl 1131
QY 2541 A-----TTCGCAAATCCTTCCAGGAGGAAA 2565

Db 1131 yIysIleTyrCysAsnProCysSerSerLys 1141

RESULT 10

S21801

myosin heavy chain, neuronal [similarity] - rat

N;Alternate names: myosin II

N;Contains: myosin ATPase (EC 3.6.4.1)

C;Species: Rattus norvegicus (Norway rat)

C;Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 19-Apr-2002

C;Accession: S21801; PN0013; S18134

R;Sun, W.; Chantler, P.D.

J. Mol. Biol. 224, 1185-1193, 1992

A;Title: Cloning of the cDNA encoding a neuronal myosin heavy chain from mammalian brain

A;Reference number: S21801; MUID:92235856; PMID:1569576

A;Accession: S21801

A;Molecule type: mRNA

A;Residues: 1-1999 <SUN>

A;Cross-references: EMBL:X62659

R;Sun, W.; Chantler, P.D.

Biochem. Biophys. Res. Commun. 175, 244-249, 1991

A;Title: A unique cellular myosin II exhibiting differential expression in the cerebral

A;Reference number: PN0013; MUID:91151356; PMID:1998509

A;Accession: PN0013

A;Molecule type: mRNA

A;Residues: 1914-1998, 'I' <SU2>

A;Experimental source: brain

A;Superfamily: myosin motor domain homology

C;Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotid

F;84-763/Domain: myosin motor domain homology <MMOT>

F;174-181/Region: nucleotide-binding motif A (P-loop)

F;541-575/Region: actin binding #status predicted

F;653-675/Region: actin binding #status predicted

F;836-1999/Domain: coiled coil #status predicted <COI>

F;836-1276/Region: S2

F;1277-1999/Region: light meromyosin

F;125/Modified site: N6,N6-trimethyllysine (Lys) #status predicted

F;180/Binding site: ATP (Lys) #status predicted

F;693,703/Active site: Cys #status predicted

F;1916/Binding site: phosphate (Ser) (covalent) #status predicted

F;1943/Binding site: phosphate (Ser) (covalent) #status predicted

Alignment Scores:

Pred. No.:	4.13e-14	Length:	1999
Score:	461.50	Matches:	210
Percent Similarity:	42.77%	Conservative:	160
Best Local Similarity:	24.28%	Mismatches:	260
Query Match:	5.97%	Indels:	235
DB:	1	Gaps:	40

US-10-788-793-1 (1-4364) x S21801 (1-1999)

QY 273 AAGACTAAGAAGTCTGTGGAGTTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATG 332

Db 832 LysValLysProLeuLeuGlnValThrArgGln----- 842

QY 333 GAAGGGGAGTTGCAGGCTCGAGAAGATGTCATCCACATGCTGAGGACAGAGAAA---ACC 389

Db 843 GluGluGluLeuGlnAlaLysAspGluGluLeu---MetLysLysValGluLysGlnThr 861

QY 390 AAGCCCGAGGTTCTGGAGGCACACTATGGATCTGCAGAACCTGAGAAAGTCTTCGG--- 446

Db 862 LysValGlu-----AlaGluLeuGluGluMetGluArgLys 873

QY 447 -----GTCCTGCACCGAGATGCCATCCTTGCTCAAGAGAAGTCCATAGGAGAAGAC 497

Db 874 HisGlnGlnLeuLeuGluGluLysAsnIleLeuAlaGluGlnLeuGlnAlaGluThrGlu 893

QY 498 GTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAAAAAGCAGAAGGAGACGTAC 557

Db 894 LeuPheAlaGluAlaGluGluMetArgAlaArgLeuAlaAlaLysLysGlnGlu----- 911

QY 558 CGCCGCATGCTAGAGCAGCTGCTG-----CTGGCTGAGAACTGTAC 599

Db 912 -----LeuGluGluIleLeuHisAspLeuGluSerArgValGluGluGluGlu 928

QY 600 AGCGCACCCGTGTACGAGCTGGAGAACGAGAAGCACAAGCACACTGACTACATGAACAAG 659

Db 929 ArgAsnGlnIle-----LeuGlnAsnGluLysLysGluGlnGlyHisLysAsnAsp 946

QY 660 AGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGAGAGAGAGGTTGAAAAAAGCTCCTTCAA 719

Db 947 LeuGluGluGlnLeuAspGluMetGluSerAlaArgGlnLysLeu----- 961

QY 720 CAAGAAAAAGCTTACCAAGCCCGCAAAAGAAAAAGGAAAAACGCTAAGCGGCTCAACAAACTT 779

Db 962 -----GlnLeuGluLysValThrThrGluAlaLys---LeuLysLysLeu 975

QY 780 CGAGATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCATGTTGGTGGACGAGAGGCAGATG 839

Db 976 GluGluGluGlnIleIleLeuGluAspGlnAsnCysLysLeuAlaLysGluLysLys--- 994

QY 840 CACATCGAGCAACTGGCCCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAAGCTGAGG 899

Db 995 -----LeuLeuGluAspArgIleAlaGluPheThrThrAsnLeuThr 1008

QY 900 GAGGAGGAAGAAAACTCAAAGCGGTCACTTACAAATCCAAGGAAGACCGCCAGAACTG 959

Db 1009 GluGluGluGluLysSerLysSerLeuAla---LysLeuLysAsnLysHisGluAlaMet 1027

QY 960 CTCAAGTTAGAGTGGACTTCGAACACAAAGGCCTCGAGGTTTTCACGAGGACGACGAAGAG 1019

Db 1028 Ile-----ThrAspLeuGluGluArgLeuArgArgGluGluLysGlnArgGlnGlu 1044

QY 1020 ATGAACGCCAAATTGGCGAATCAAGAATCTCACAAACCGGCAACTTCGACTCAAACCTGGTT 1079

Db 1045 Leu-----GluLysThrArgArgLysLeuGluGlyAspSerThr 1057

QY 1080 GGCTTATCGCAAAGGATTGAGGAGCTGGAA-----GAGACCAATAAAACCTT 1127

Db 1058 AspLeuSerAspGlnIleAlaGluLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeu 1077

QY 1128 CAGAAGGCAGAGGAGAGCTCCAGGAGCTGAGAGAGAAAATTCCCAAGGGGAATCTGTGGA 1187

Db 1078 AlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgValGluGluAlaAlaGln 1097

QY 1188 AACTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTGTGTGAGATGGAG--- 1244

Db 1098 LysAsnMetAlaLeuLysLysIleArgGluLeuGluSerGlnIleSerGluLeuGlnGlu 1117

QY 1245 -----GGC 1247

Db 1118 AspLeuGluSerGluArgAlaSerArgAsnLysAlaGluLysGlnLysArgAspLeuGly 1137

QY 1248 AAGGATGAAGAGATCACGAAGACCGAG-----GCCAGTGC 1283

Db 1138 GluGluLeuGluAlaLeuLysThrGluLeuGluAspLeuThrAspSerThrAlaAlaGln 1157

QY 1284 CGGGAGCTGAAG-----AAGAAGCTCCAAGAGGAA 1313

Db 1158 GlnGluLeuArgSerLysArgGluGlnGluValAsnIleLeuLysLysThrLeuGluGlu 1177

QY 1314 GAACACCACAGCAAGGAACCTTAGACTAGAACTGGAGAGAGCTGCAGAAGAGGATGTCTGAG 1373

Db 1178 GluAlaLysThrHisGluAlaGlnIleGln---GluMetArgGlnLysHisSerGlnAla 1196

QY 1374 CTGGAGAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTG 1433

Db 1197 ValGluGluLeuAlaGluGlnLeuGluGlnThrLysArgLysValAla----- 1212

QY 1434 AACCTGGAGAAGGGAAG-----AACCTAACCAAGACCTGCTGAACGAGCTG 1481

Db 1213 AsnLeuGluLysAlaLysGlnThrLeuGluAsnGluArgGlyGluLeuAlaAsnGluVal 1232

QY 1482 GAGGTG-----GTCAAGAGTCTGA 1499

Db 1233 LysValLeuLeuGlnGlyGlyArgAspSerGluHisLysArgLysLysValGluAlaGln 1252

QY 491 AGAAGACGCTCTATGAGAAACCTATCTCAGAGCTGGACAGA-----CTGGAGGAAACGA 544
Db 911 -----AspPheGluSerGlnIleLysGluLeuGluArgLeuLeuAspGluGluAs 928
QY 545 GAAGGAGACGTACCGCCGCATGCTAGAGCAGCTGCTGCTGGCTGAGAAGTGTAC---AG 601
Db 928 pAlaAlaAlaAspLeuGluGlyIleLysLysLysMetGluAlaAspAsnAlaAsnLeuLy 948
QY 602 GCGCACCGTGTACGAGCTGGAGAACGAGAACACAAG---CACACTGACTACATGAACAA 658
Db 948 sLysAspIleGlyAspLeuGluAsnThrLeuGlnLysAlaGluGlnAspLysAlaHisLy 968
QY 659 GAGCGACGACTTCACCAACCTG-----CTGGAGCAGGAGCGAGAGGTTGAAAAA 709
Db 968 sAspAsnGlnIleSerThrLeuGlnGlyGluIleSerGlnGlnAspGluHisIleGlyLy 988
QY 710 GCTCCTTGAACAAGAAAGTTTACCAAGCCCGCAAGAAAAGGAAACGCTAAGCGGCT 769
Db 988 sLeuAsnLysGluLysLysAlaLeu-----GluGluAlaAsnLysLysThrSe 1004
QY 770 CAACAAACTTCGAGATGAG-----CTTGTGAAGCTCAAGTCTTCGCG 811
Db 1004 rAspSerLeuGlnAlaGluGluAspLysCysAsnHisLeuAsnLysLeuLysAlaLysLe 1024
QY 812 CCTCATGTTGGTGACGAGGCGCAGATGCACATCGAGCAACTGGGCTGCAGAGTCAGAA 871
Db 1024 uGluGlnAlaLeuAspGluLeuGluAspAsnLeuGlu-----ArgGluLysLy 1040
QY 872 AGTCCAG---GACCTCACTCAGAAAGCTGAGGGAGGAGGAAGAAAAAATCTCAAAGCGGTAC 928
Db 1040 sValArgGlyAspValGluLysAlaLysArgLysValGluGlnAspLeuLysSer---- 1058
QY 929 TTACAAATCCAAGGAAGACCGCCAGAAAGCTGCTCAAGTTAGAAGTGGACTTCGAACAA 988
Db 1059 -----ThrGlnGluAsnValGluAspLeuGluArgValLysArgGluLeuGluGluAs 1076
QY 989 GGCCTCGAGGTTTTCCAGGAGCAGCAAGAGATGAACGCCAAATTTGGCGAATCAAGAAATC 1048
Db 1076 nValArgArgLysGluAlaGluIleSerSerLeuAsnSerLysLeuGluAspGluGlnAs 1096
QY 1049 TCACAACCGGCAACTTCGACTCAAACTGGTTGGCTTATCGCAAGGATTGAGGAGCTGGA 1108
Db 1096 nLeuValSerGlnLeuGlnArgLysIleLysGluLeuGlnAlaArgIleGluGluLeuGl 1116
QY 1109 AGAG-----ACCAATAAAGCCCTTCAGAAGGC 1135
Db 1116 uGluGluLeuGluAlaGluArgAsnAlaArgAlaLysValGluLysGlnArgAlaGluLe 1136
QY 1136 AGAGGAAGAGCTCCAGGAGCTGAGAGAGAAATTTGCCAAAGGGGAATGTGGAAACTCCAG 1195
Db 1136 uAsnArgGluLeuGluGluLeuGlyGluArgLeu-----AspGluAlaGlyGlyAlaTh 1154
QY 1196 TCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGGCTGCTTGAGATG-----GA 1243
Db 1154 r---SerAlaGlnIleGluLeuAsnLysLysArgGluAlaGluLeuLeuLysIleArgAr 1173
QY 1244 GGGCAAGGATGAAGAGATCACGAAGACCGAGGCCCACTGCGGGAGCTGAAGAAGAAGCT 1303
Db 1173 gAspLeuGluGluAlaSerLeuGlnHisGluAlaGlnIleSerAlaLeuArgLysLysHi 1193
QY 1304 CCAAGAGGAAGAACACCACAGCAAGGAACTTAGACTAGAAGTGGAGAAGCTGCAGAAGAG 1363
Db 1193 sGlnAspAla-----AlaAsnGluMetAlaAspGlnValAspGlnLeuGlnLysVa 1210
QY 1364 GATGCTGTGAGCTGGAGAAG-----1382
Db 1210 lLysSerLysLeuGluLysAspLysLysAspLysArgGluMetAspAspLeuGluSe 1230
QY 1383 -----CTGGAGGAAGCGTTTCAGCCGGAGTAA 1408
Db 1230 rGlnMetThrHisAsnMetLysAsnLysGlyCysSerGluLysValMetLysGlnPheGl 1250
QY 1409 GTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAGGAGAAACCTTAACCAAGACCT 1468

Db 1250 uSerGlnMetSerAspLeuAsnAlaArgLeuGluAspSerGlnArg-----Se 1266
QY 1469 GCTGAACGAGCTGGAGGTGGTCAAGAGTCCGAGTT-----AA 1504
Db 1266 rIleAsnGluLeuGlnSerGlnLysSerArgLeuGlnAlaGluAsnSerAspLeuThrAr 1286
QY 1505 AGAACTCGAATGCTCCGAGAGTAGACTG-----GAGAAGGCCGAGTTAAG 1549
Db 1286 gGlnLeuGluAspAlaGluHisArgValSerValLeuSerLysGluLysSerGlnLeuSe 1306
QY 1550 C-----CTCAAAGATGACCTTACAAAGCTGAAGTCTTCACTGTGATGCTGTGGATGA 1603
Db 1306 rSerGlnLeuGluAspAlaArgArgSerLeuGluGluThr-----ArgAl 1322
QY 1604 GAGGAAAAATATGATGGAGAAAAATAAAGCAAGAGAGAGGAAAAAGTGGATGGTTGAATAA 1663
Db 1322 aArgSerLysLeuGlnAsnGluValArgAsnMetHisAlaAspMetAspAlaIleArgGl 1342
QY 1664 AAACCTTTAAGGTGGAGCGGAAAAAGTCAATGATGTGACGGAAGAAAGCTAATCGAGGAAAG 1723
Db 1342 uGlnLeuGluGluGlnGluSerLysSerAspValGlnArgGlnLeuSerLysAlaAs 1362
QY 1724 CAAGAAGCTTTTAAAACTCAAATCTGAAAATGGAGAAAAAGGAGTACAGTCTGCACAAAGGA 1783
Db 1362 nAsnGluIleGlnGlnTrpArgSerLysPheGluSerGluGlyAlaAsnArgThrGluGl 1382
QY 1784 GAGGGATGAG-----CTGATGGGTAACTGAGGAGCGGAAGAAAGTCTCTG 1831
Db 1382 uLeuGluAspGlnLysArgLysLeuLeuGlyLysLeu---SerGluAlaGluGlnThrTh 1401
QY 1832 TGAACGTGAGC-----TGAGT---GTAGACTTACTAAAAGACGGCTT---GATGG 1876
Db 1401 rGluAlaAlaAsnAlaLysCysSerAlaLeuGluLysAlaLysSerArgLeuGlnGlnGl 1421
QY 1877 CATAGAGGAGGTAGAAAAGGAAATAAACCGAGGTAGGTCTGCAAGGGG----- 1925
Db 1421 uLeuGluAspMetSerIleGluValAspArgAlaAsnAlaSerValAsnGlnMetGluLy 1441
QY 1926 -----TCTGAGTTCACTGCTGCCGGAAGACAAATAAGATCAGAGAACTAAC 1969
Db 1441 sLysGlnArgAlaPheAspLysThrThrAlaGluTrpGlnAlaLysValAsnSerLeuGl 1461
QY 1970 GCTTGAAATCGAGAGACTGAAGAAACGGCTCCAG-----CAGTTGAGGTGGT 2017
Db 1461 nSerGluLeuGluAsnSerGlnLysGluSerArgGlyTyrSerAlaGluLeuTyrArgIl 1481
QY 2018 GGAGGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAGCAGAAAGTTCAGAAC 2077
Db 1481 eLysAlaSerIleGluGluTyrGlnAspSerIleGlyAlaLeuArgArgGluAsnLysAs 1501
QY 2078 CGAGCAGGATAAGGCCAAACTTCTCTCTCCAGCAGCTCGAGGAA-----ATCAAAACA 2128
Db 1501 nLeuAlaAspGluIleHisAspLeuThrAspGlnLeuSerGluGlyArgSerThrHi 1521
QY 2129 CCAAATGGCCCAAGCACAAA-----GCCATAGAGAAAGGGGAGGCCGTGAGCCAGGA 2179
Db 1521 sGluLeuAspLysAlaArgArgArgLeuGluMetGluLysGluGluLeuGlnAlaLe 1541
QY 2180 AGCCGAACCTGCGACACAGGTTTCGGCTGGAGGAGGCTAAAAGTCGTGATTACAGGCCGA 2239
Db 1541 uGluGluAlaGluGlyAlaLeuGluGlnGluAlaLysValMetArgAlaGlnLeuGl 1561
QY 2240 GGTGCAGGCTCTCAAGGAGAAAGATCCACAG---CTGATGAACAAGGAAGACCAGCTGTC 2296
Db 1561 uIleAlaThrValArgAsnGluIleAspLysArgIleGlnGluLysGluGluPheAs 1581
QY 2297 TCAGCTCCAAAGTCGACTATTTCG-----GTCCTTCAGCAAAAGATTATGGAAGAAGA 2347
Db 1581 pAsnThrArgArgAsnHisGlnArgAlaLeuGluSerMetGlnAlaSerLeuGluAlaGl 1601
QY 2348 AACTAAGAACAAACATGCGGAGGGAGGTCTCTCAATCTGACCAAGGAGCTAGAGCTTTC 2407

Db	1601	uAlaLysGlyLys-----AlaAspAlaMetArgIleLysLysLysLeuGlnAs	1618
QY	2408	C---AAGCGTACAGCCGAGCTCTCAGGCCGAGTGGAAACGGCCGAAGGTGGACGT	2464
Db	1618	pileAsnGluLeuGluValAlaLeuAspAlaSerAsnArgGlyLysAlaGluMetGluLy	1638
QY	2465	GCCTGTGGCCTCCACTGGGTGCAGACCGAGCGGTGTGCGGGGATGTCGCGGAGGGA	2524
Db	1638	sThrValLysArgTyrGlnGlnGlnIleArgGluMetGlnThrSerIleGluGluGluG	1658
QY	2525	GACCCCGGTGTTCATTCCGAAATCCTTCCAG-----	2558
Db	1658	nArgGlnArgAspGluAlaArgGluSerTyrAsnMetAlaGluArgArgCysThrLeuMe	1678
QY	2559	-----GAGGAAATCACATCATGAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCAT	2611
Db	1678	tSerGlyGluValGluGluLeuArgAlaAlaLeuGluGlnAlaGluArgAlaArgLysAl	1698
QY	2612	GGAACGGTCTCGGTCTCTCGACAGGTATCCCCCAGCAGCGGAATGAGCTCACCATGAGGAA	2671
Db	1698	aSerAspAsnGluLeuAlaAspAla---AsnAspArgValAsnGluLeuThrSerGlnVa	1717
QY	2672	GTCTTGGATTCTTGGATGAGAAAAAGAGAAAACGGTCT-----TCCACTCC	2719
Db	1717	lSerSerValGlnGlyGlnLysArgLysLeuGluGlyAspIleAsnAlaMetGlnThrAs	1737
QY	2720	GCAGGAGAAAGGCCAGGCCAAACCCAGGGTGCAGGGCCACCCCGGGAGTGGTCCTAGC	2779
Db	1737	pLeuAspGluMetHisGlyGluLeuLysGlyAlaAspGluArgCysLysLysAlaMetAl	1757
QY	2780	ACCAAAGCGGCCAGCCCTACATCCGTGTGACACCAGATCATGAGAAC	2831
Db	1757	aSpAlaAlaArgLeuAlaAspGluLeuArgAlaGluGlnAspHisserAsn	1774
RESULT 12			
A47297			
myosin heavy chain form B, nonmuscle - African clawed frog			
C;Species: Xenopus laevis (African clawed frog)			
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 02-Feb-2001			
C;Accession: A47297; A55441			
R;Bhatia-Dey, N.; Adelstein, R.S.; Dawid, I.B.			
Proc. Natl. Acad. Sci. U.S.A. 90, 2856-2859, 1993			
A;Title: Cloning of the cDNA encoding a myosin heavy chain B isoform of Xenopus nonmuscl			
A;Reference number: A47297; MUID:93219383; PMID:8464900			
A;Accession: A47297			
A;Status: preliminary; not compared with conceptual translation			
A;Molecule type: mRNA			
A;Residues: 1-1992 <BHA>			
A;Cross-references: GB:L09740; NID:g214623; PIDN:AAA49915.1; PID:g214624			
A;Experimental source: XTC cells			
A;Note: sequence extracted from NCBI backbone (NCBIP:128722)			
R;Kelley, C.A.; Oberman, F.; Yisraeli, J.K.; Adelstein, R.S.			
J. Biol. Chem. 270, 1395-1401, 1995			
A;Title: A Xenopus nonmuscle myosin heavy chain isoform is phosphorylated by cyclin-p34			
A;Reference number: A55441; MUID:95138137; PMID:7836406			
A;Accession: A55441			
A;Status: preliminary			
A;Molecule type: protein			
A;Residues: 198-232 <KEL>			
C;Superfamily: myosin heavy chain; myosin motor domain homology			
C;Keywords: nucleotide binding; P-loop			
F;88-787/Domain: myosin motor domain homology <MMOT>			
F;178-185/Region: nucleotide-binding motif A (P-loop)			
Alignment Scores:			
Pred. No.:	9.65e-14	Length:	1992
Score:	453.50	Matches:	206
Percent Similarity:	43.98%	Conservative:	185
Best Local Similarity:	23.17%	Mismatches:	346
Query Match:	5.87%	Indels:	152
DB:	2	Gaps:	33
US-10-788-793-1 (1-4364) x A47297 (1-1992)			

QY	101	AAGTTCATCTAACGGGCATGTCTCTCTGCCCAAGTCTCTCATCATCAGCAGTGTGGTG	160
Db	833	LysValLeuGlnArgAsnCysAlaAlaTyrLeuLysLeuArgHisTrpGln---TrpTrp	851
QY	161	TAAGGGCCCTCAGAAGATGCAAAAAAGAACCAATCGGAAGGAGGAGGATGTTCAT	220
Db	852	-ArgLeuPheThrLysValLysProLeuGlnValThrArgGlnGluGluLeuVa	871
QY	221	GGCTTCCGGAACCTATCAAAGGCACCTCAAACCATCTGAGAAAGTGAGAAAAAGACTAA	280
Db	871	lAlaLysAspGluGluLeuLysValLys-----GluLysGlnSerLy	886
QY	281	GAAGTCTGTGGAGTTA-----TCCAAGGAGGACCTCATCCAGCTCCTGAG	325
Db	886	sValGluGlyGluLeuValAspMetGluGlnLysHisGlnGlnLeuValGluGluLysAs	906
QY	326	TATCATGGAAGGGAGTTGCAGGCTCGAGAAGATGTCTCCACATGCTGAGGACAGAGAA	385
Db	906	nIleLeuAlaGluGlnLeuHisAlaGluThrGluLeuPhe-----	919
QY	386	AACCAAGCCCGAGTTCTGGAGGCACACTATTGA-----TCTGCAGAACCTGAGAAAGT	439
Db	920	-AlaGluAlaGluMetArgAlaArgLeuAlaIleLysLysGlnGluMetGluGluI	939
QY	440	GCTTCGGGTCTGCACCGAGATGCCATCCTTGTCTCAAGAGAAGTCCATAGGAGAAGACGT	499
Db	939	eLeuArgAspLeuGlu-----IleArgMetGluGluGluGluArgAsnGlnVa	956
QY	500	CTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAAAAAGCAGAGGAGACGTACCG	559
Db	956	lLeuGlnAsn-----GluLysLysLysMetGlnThrHisVa	968
QY	560	CCGCATGCTAGACAGCTGCTGGCTGAGAAGTGTCAACAGGCGCACCGTGTACGAGCT	619
Db	968	lGlnAspLeuGluGlnLeuAspGluGluAlaAlaGlnLys-----Le	984
QY	620	GGAGAAGCAGAAAGCAACAGCACACTGACTACATGAACAAGAGCGACGACTTCACCAACT	679
Db	984	uGlnLeuGluLysValThrAlaGluAlaLysIleLysLysMetGluGluAspIleLeuVa	1004
QY	680	GCTGGAGCAGGAGCGAGAGAGGTTG-----AAAAAGCTCCTTGAACAA-----	722
Db	1004	lLeuGluAspGlnAsnSerLysPheLeuLysGluLysLysLeuLeuGluArgIleAl	1024
QY	723	-GAAAAAGCTTACCAAGCCCGCAAGAAAAAGAAACGCTAAGCGGCTCAACAAACTTCG	781
Db	1024	aGluSerThrSerGlnLeuAlaGluGluGluLysAlaLysAsnLeuAlaLysLeuLy	1044
QY	782	AGATGAGCTTGTGAAGCTCAAGTCTCTCGCCCTCATGTTGGTGGACGAGAGGCAGATGCA	841
Db	1044	s-----AsnLysGlnGluMetMe	1050
QY	842	CATCAGCAACTGGGCTGCAGAGTCAGAAAAGTCCAGGACCTCACTCAGAAGCTGAGGGA	901
Db	1050	tIleSerAspLeuGluGluArgLeuLysLysGluGluLysThrArgGlnGluLeuGluLy	1070
QY	902	GGAGGAAGAAAACTCAAAAGCGGTCACTTACAAATCCAAGGAAGACCGCCAGAAAGTCT	961
Db	1070	sAlaLysArgLysLeuAspGlyGluThr-----ThrAspPheGlnAspGlnIleAl	1087
QY	962	CAAGTTAGAAGTGGACTTCGAACACAAAGGCCTCGAGGTTTTCGCCAGGAGCACGAAGAGAT	1021
Db	1087	aGluLeuGlnAlaGlnIleGluGluLeuLysLeuGlnLeuAlaLysLysGluGluGluLe	1107
QY	1022	GAACGCCAAATTTGGCG-----AATCAAGAATCTCACAAACCGGCAACTTCGACTCAAAC	1075
Db	1107	uGlnAlaAlaLeuAlaArgGlyAspGluGluValLeuGlnLysAsnAsnThrLeuLysLe	1127
QY	1076	GGTT-----GGCTTATCGCAAAAGGATTGAGGAGCTGGAAGAGACCAATAA	1120
Db	1127	uValArgGluLeuGlnAlaGlnIleAlaGluLeuGlnGluAspLeuGluSerGluLysAl	1147

QY	1121	AAGCCTTCAGAAGGCAGAGAAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAGGGGA	1181
Dd	1147	aSerArgAsnLysAlaGluLysGlnLysArgAspLeuSerGluGluLeu-----	1163
QY	1181	ATGTGGAAACTCCAGTCTCATGGCGGAAGTGAGAGTCTGCGCAAGCGCTGCTTGAGAT	1240
Dd	1164	-----GluAlaLeuLysThrGluLeuGluAap-----ThrLeuAspTh	1176
QY	1241	GGAGGGCAAGGATGAAGATC---ACGAAGACCAGGCCCCAGTGCCGGGAGCTGAAGAA	1297
Dd	1176	rThrAlaAlaGlnGlnGluLeuArgThrLysArgGluGlnGluValAlaGluLeuArgLy	1196
QY	1298	GAAGCTCCAAGAGGAA--GAACACCACAGCAAGGAACCTTAGACTAGAAAGTGAGAGAAGCT	1354
Dd	1196	sSerIleGluGluGluThrArgAsnHisGluAlaGlnIleGln-----GluMetAr	1213
QY	1355	GCAGAAGAGGATGTTCTGAGCTGGAGAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGA	1414
Dd	1213	gGlnArgGlnAlaThrAlaLeuGluGluLeuSerGluGlnLeuGluGlnAlaLys----	1233
QY	1415	ATGCACCCAGCTCCATCTGAACCTGGAGAAGGAGGAACCTA-----ACCAA	1462
Dd	1232	-----ArgPheLysValAsnLeuGluLysAsnLysGlnSerLeuGluSerAspAsnLy	1249
QY	1463	AGACCTGCTGAACAGCTGGAGGTGGTCAAGAGTCCAGTCTTAAGAACAATCGAATGCTCCGA	1522
Dd	1249	sGluLeuAlaThrGluValLysSerLeuGlnGlnMetLysAlaGluSerGluTyrLysAr	1269
QY	1523	GAGTAGACTGGAG-----AAGGCCGAGTTA-----AGCCT	1552
Dd	1269	gLysLysLeuGluGlyGlnValGlnGluLeuHisAlaLysValLeuGluGlyAspArgLe	1289
QY	1553	CAAAGATGACCTTACAAAAGCTGAAGTCTCTTCACTGTGTGATGCTGGTGGATGACAGGAAAA	1612
Dd	1289	uArgAlaAspMetValGluLysSerSer-----LysLeuGlnAsnGluLeuGluAs	1306
QY	1613	TATGATGGAGAAAAATAAACCAAGAGAGAGGAAAGTGGATGGTTGAATAAAAACTTTAA	1672
Dd	1306	nValSerSerLeuLeuGluGluAlaGluLysGlyIleLysLeuAlaLysAspValAl	1326
QY	1673	GGTGGAGCAGGAAAAAGTCAATGGATGTGACGGAAAAAGCTAATCGAGGAAAGCAAGAAGCT	1732
Dd	1326	aSerMetGluSerGlnLeuGlnAspThrGlnGluLeuGlnGluThrArgGlnLy	1346
QY	1733	TTTTAAACTCAATCT-----GAAATGGAGGAAAAAGGAGTACAGTCTGACAAAGGA	1783
Dd	1346	sLeuAsnGlnSerSerArgIleArgGlnLeuGluGluLysAsnAsnLeuGlnGluGl	1366
QY	1784	GAGGGATGAGCTGATGGTTAACTGAGGAGCGAAGAAAGGTCCTGTGAACCTGAGCTG	1843
Dd	1366	nGlnGluGluGluGluAlaArgLysSerLeuGluLysGlnIleLeuSerLeuGlnSe	1386
QY	1844	CAGTGTAGACTTACTAAAGAAGCGGCTTGAT-----GGCATAGA	1882
Dd	1386	rGlnLeuIleGluAlaLysLysLysValAspAspGluValGlyThrIleGluGlyLeuGl	1406
QY	1883	GGAGGTAGAAAGGGAATAAACCGAGGTAGTCTGTCGAAGGGTCTGAGTTCACCTGCCC	1942
Dd	1406	uGluValLysLysLysLeuLysAspThrGluGlyLeuGly-----	1420
QY	1943	GGAAGACAATAAGATCAGAGAACTAACGCTTGAATCGAGAGACTGAAGAAACGGCTCCA	2002
Dd	1421	-----GlnArgLeuGluGluLysIleIleAlaTyrgLuLysLeuGluLysThrLysAs	1438
QY	2003	GCAGTTGGAGGTGGTGGAGGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGA	2062
Dd	1438	nArgLeuGlnGlnGluLeuAspAspLeuMetValAspLeuAspHisGlnArgGlnIleVa	1458
QY	2063	GCAGAAGTTCAGAACCGAGCAGGATAAGGCAAACTTCCTCTCCAGCAGCTCGAGGAAAT	2122
Dd	1458	lSerAsnLeuGluLysLysGlnLysLys-----PheAspGlnLeuLeuAlaGluGl	1475
QY	2123	CAAAACCAAAATGGCCAAGCAC-----	2144

```

Db      1475 uLysAsnIleSerAlaArgHisAlaGluGluArgAspArgAlaGluAlaAspAlaArgGl 1499
QY      2145 -----AAAGCCATAGAGAAAGGGGACGCCGTGAGCCAGGAAGCCGAACCTGCAGACA 2194
Db      1495 uLysGluThrLysAlaLeuSerLeuAlaArgAlaLeuAaspGluAlaLeuGluAlaGlnAs 1515
QY      2195 CAGGTTT---CGGGTGGAGGAG-----GCTAAAAGTCGTGATTATCAGGCCGAGGT 2242
Db      1515 pGluPheGluArgLeuAsnLysGlnLeuArgAlaGluMetGluAspLeuMetSerSerLy 1535
QY      2243 GCAGGCTCTCAAGGAGAAGATCCACGAGCTGATGAACAAGAACACCAGCTGTCTCAGCT 2302
Db      1535 sAspAspValGlyLysAsnValHisGluLeu-----GluLysSerLysAr 1550
QY      2303 CCAAGTCGACTATTCGGTCCTTCAGCAAAGATTTATGGAAGAAGAACTAAGAACAAAGAA 2362
Db      1550 gAlaLeuAspGlnGlnValGluGluMetArgThrGlnLeuGluGluLeuGluAspGluLe 1570
QY      2363 CATGGGGAGGAGGTCTCAATCTGACCAAGAGCTAGAG-----CTTTCCAAGCG 2413
Db      1570 uGlnGlyThrGluAspAlaLysLeuArgLeuGluValAsnMetGlnAlaMetLysAlaGl 1590
QY      2414 CTACAGCCGAGCTCTCAGCCGAGTGGAACGGCCGAAGGATGGTGGACCTGCCTGTGGC 2473
Db      1590 nPheGluArgAspLeuGlnThrArgAspGluGlnAsnGluGluLysLysArgAlaLeuVa 1610
QY      2474 CTCCTACTGGGGTGCAGACCGAGGCGGTGTCGGGGGATGCTCGGAGGAGGAGACCCCGC 2533
Db      1610 lLysGlnValArgGluLeuGluAlaGluLeuGluAspGluArgLysGlnArgAlaMetAl 1630
QY      2534 TGTGTTCAATTCGCAAAATCCTTCCAG 2558
Db      1630 aValAlaIleLysLysLysLeuGlu 1638

RESULT 13
A43336
microtubule-vesicle linker CLIP-170 - human
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 05-Nov-1999
R;Accession: A43336
R;Pierre, P.; Scheel, J.; Rickard, J.E.; Kreis, T.E.
Cell 70, 887-900, 1992
A;Title: CLIP-170 links endocytic vesicles to microtubules.
A;Reference number: A43336; MUID:92405160; PMID:1356075
A;Accession: A43336
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1392 <PIE>
A;Cross-references: GB:M97501; NID:g180621; PIDN:AAA35693.1; PID:g180622

Alignment Scores:
Pred. No.:          1.27e-13          Length:          1392
Score:              450.50           Matches:         202
Percent Similarity: 43.65%           Conservative:    183
Best Local Similarity: 22.90%        Mismatches:     316
Query Match:       5.83%             Indels:         181
DB:                 2                Gaps:           38

US-10-788-793-1 (1-4364) x A43336 (1-1392)
QY      162 AAGGGCCCTCAGAAGATGCA-----AAAAAGAACAAGGCCAATCGGAAGGAGGAG 212
Db      452 LysGlyAspLeuGluValAlaThrValSerGluLysSerArgIleMetGluLeuGluLys 471
QY      213 GAT-----GTCATGGCTTCGGAACTATCAAAAGGCCACTC-----AAA 251
Db      472 AspLeuAlaLeuArgValGlnGluValAlaGluLeuArgArgLeuGluSerAsnLys 491
QY      252 CCATCTGGAGAAAGTGAGAAAAAGACTAAGAAGTCTGTGGAGTTATCCAAGGAGGACCTC 311
Db      492 ProAlaGlyAsp-----ValAspMetSer-----ValAspMetSer----- 499

```


Db 1177 -----AspAlaAspGluGluLysAlaSerLeuGlnLysSerIleSerIleThrSerAla 1194

QY 2217 AAAAGTCGTGATTACAGGCCGAGGTGCAGGCTCTCAAGGAGAAGATCCACGAGCTGATG 2276

Db 1195 LeuLeuThrGluLysAspAlaGluLeuGluLysLeuArgAsnGluValThrValLeuArg 1214

QY 2277 AACAAAGGAAGACCAGCTGTCTCAGCTCCAAAGTCGACTATTCCGGTCCTTCAGCAAAGATTT 2336

Db 1215 GlyGluAsnAlaSerAlaLysSerLeuHisSerValValGlnThrLeuGluSerAspLys 1234

QY 2337 ATGGAAGAAGAACTAAGAACAAAGAACATGGGAGGAGGTCTCAATCTGACCAAGGAG 2396

Db 1235 ValLysLeuGluLysValLysValLysAsnLeu-----Glu 1245

QY 2397 CTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGGAAGGATG 2456

Db 1246 LeuGlnLeuLysGluAsnLysArgGlnLeuSerSerSerSerGlyAsnThrAspThrGln 1265

QY 2457 GTGGAC 2462

Db 1266 AlaAsp 1267

RESULT 14

JC5421

smooth muscle myosin heavy chain 2 - mouse

C;Species: Mus musculus (house mouse)

C;Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 02-Feb-2001

C;Accession: JC5421

R;Hasegawa, K.; Arakawa, E.; Oda, S.; Matsuda, Y.

Biochem. Biophys. Res. Commun. 232, 313-316, 1997

A;Title: Molecular cloning and expression of murine smooth muscle myosin heavy chains.

A;Reference number: JC5420; MUID:97242182; PMID:9125171

A;Accession: JC5421

A;Molecule type: mRNA

A;Residues: 1-1938 <HAS>

A;Cross-references: DDBJ:D85924; NID:g1945079; PIDN:BAA19691.1; PID:g1945080

A;Experimental source: smooth muscle

C;Comment: This protein plays a role in smooth muscle cell contraction.

C;Superfamily: myosin heavy chain; myosin motor domain homology

C;Keywords: nucleotide binding; P-loop

F;88-771/Domain: myosin motor domain homology <MMOT>

F;178-185/Region: nucleotide-binding motif A (P-loop)

Alignment Scores:

Pred. No.:	1.55e-13	Length:	1938
Score:	449.00	Matches:	203
Percent Similarity:	42.82%	Conservative:	167
Best Local Similarity:	23.50%	Mismatches:	292
Query Match:	5.81%	Indels:	202
DB:	2	Gaps:	33

US-10-788-793-1 (1-4364) x JC5421 (1-1938)

QY 276 ACTAAGAAGTCTGTGGAGTTATCCAAGGAGGACCTCATCCAG----- 317

Db 807 ThrLysArgGlnGlnLeuThrAlaMetLysValIleGlnArgAsnCysAlaAlaTyr 826

QY 318 ----- 326

Db 827 LeuLysLeuArgAsnTrpGlnTrpTrpArgLeuPheThrLysValLysProLeuLeuGln 846

QY 327 ATC-----ATGGAAGGGAGTTGCAGGCTCGAGAAGATGTCATCCACATGCTG----- 374

Db 847 ValThrArgGlnGluGluMetGlnAlaLysGluGluMetGlnLysIleThrGlu 866

QY 375 AGGACAGAGAAAAACCAAGCCGAGGTTCTGGAGGCACACTATGGATCTGCAGAACCTGAG 434

Db 867 ArgGlnLysAlaGluThrGluLeuLysGluLeuGlu----- 879

QY 435 AAAGTGCTTCGGGTCTCGACCGAGATGCCATCTTCCTCAAGAGAAGTCCATAGAGAA 494

Db 880 -----GlnLysHisThrGlnLeuAlaGluLysThrLeu----- 891

QY 495 GACGTCTATGAGAAACCTATCTCAGAGCTGGACACACTGGAGGAAAAAGCAGAAGCAGACG 554

Db 892 ---LeuGlnGluGlnLeuAlaGluThrGluLeuTyrAlaGluSerGluGluMetArg 910

QY 555 TACCGCCGCTAGACGAGCTGCTGCTGGCTGAGAAAGTGTACAGGGCCACCGTGTAC 614

Db 911 ValArg-----LeuAlaAlaLysLys-----Gln 918

QY 615 GAGCTGGAGAACGAGAAGCACAAAGCACACTGACTACATGAAACAAGAGCGACGACTTCACC 674

Db 919 GluLeuGluGluIleLeuHisGluMetGluAlaArgLeuGluGluGluAlaAspArgArg 938

QY 675 AACCTGCTGGAGCAGGAGCGAGAGAG-----TTGAAAAAAGCTC 713

Db 939 GlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGlnMetLeuAspLeuGluGluGln 958

QY 714 CTTGAACAAGAAAAAGCTTACCAAGCCCGCAAGAAAAAGAAACGCT-----AAG 764

Db 959 LeuGluGluGluAlaAlaArgGlnLysLeuGlnLeuGluLysValThrAlaGluAla 978

QY 765 CGGCTCAACAACTTCGAGATGAGCTTGTGAAGCTCAAGTCTTCGCCCTCATGTTGGTG 824

Db 979 LysIleLysLysLeuGluAspIle-----LeuValMetAsp 991

QY 825 GACGAGAGGCAGATGCACATCGAGCAACTGGGCTGCAGAGTCAGAAAGTCCAGGACCTC 884

Db 992 AspGlnAsnSerLysLeuSerLysGluArgLysLeuLeuGluGluArgValSerAspLeu 1011

QY 885 ACTCAGAAAGCTGAGGAGGAGGAAGAAAAAACTCAAAAGCGGTCACT----- 929

Db 1012 ThrThrAsnLeuAlaGluGluGluLysAlaLysAsnLeuThrLysLeuLysSerLys 1031

QY 930 -----TACAAATCCAAGGAAGACCGCCAG 953

Db 1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

QY 954 AAGCTGCTCAAGTTAGAAGTGGACTTCGAACACAAAGGCCTCGAGGTTTCCAGGAGCAC 1013

Db 1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071

QY 1014 GAAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCACAAACCGGCAACTTCGACTCAA 1073

Db 1072 AlaAspLeuGlnAlaGlnIleAla-----GluLeuLysMetGln 1084

QY 1074 CTG-----GTTGGCTTATCGAAAGGATTGAGGAGCTGGAA 1109

Db 1085 LeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgLeuAspGluGluIleAla 1104

QY 1110 GAGACCAATAAAAGCCTTCAGAAG-----GCAGAG 1139

Db 1105 GlnLysAsnAsnAlaLeuLysLysIleArgGluLeuGluGlyHisIleSerAspLeuGln 1124

QY 1140 GAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAAACTCCAGTCTC 1199

Db 1125 GluAspLeuAspSerGluArgAlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeu 1144

QY 1200 ATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTG-----CTTGAGATGGAGGCAAG 1250

Db 1145 GlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeuAspSerThrAlaThr 1164

QY 1251 GATGAAGAGATC---ACGAAGACCCGAGGCCAGCTGCGGGAGCTGAAGAAGAAGCTCCAA 1307

Db 1165 GlnGlnGluLeuArgAlaLysArgGluGlnGluValThrValLeuLysLysAlaLeuAsp 1184

QY 1308 GAGGAA--GAACACCACAGCAAGGAACCTTAGACTAGAAGTGGAGAAGCTGCAGAAAGAG 1364

Db 1185 GluGluThrArgSerHisGluAlaGlnValGlnGluMetArgGlnLysHisThrGlnAla 1204

QY 1365 ATGTCTGAGCTGGAAGAGCTGGAGGAAGCGTTTCAGCCCGAGTAAGTCGGGAATGCACCCAG 1424

Db 1205 ValGluGluLeuThrGluGlnLeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLys 1224

QY 1425 CTCCATCTGAACCTGGAGAAGGAGAAGAACCTAACCAAAAGACCTGTCTGAACGAGCTGGAG 1484

Db 1225 SerLysGlnThrLeuGluLysGlu-|||||-----AsnAlaAspLeuAlaGlyGluLeuArg 1241
QY 1485 GTG-|||||-----GTCAAGAGTCGAGTTAAAGAACTCGAA----- 1514
Db 1242 ValLeuGlyGlnAlaLysGlnGluValGluHisLysLysLysLeuGluValGlnLeu 1261
QY 1515 -----TGCTCCGAGAGTAGACTGGAGAGGCCGAGTTAAGCCTCAAA 1556
Db 1262 GlnAspLeuGlnSerLysCysSerAspGlyuArgAlaGluLeuSer----- 1279
QY 1557 GATGACCTTACAAAGCTGAAGTCCTTCACTGTGATGCTGGTGATGAGAGGAAAAATATG 1616
Db 1280 AspLysValHisLysLeuGln-|||||:-----AsnGluValGluSerVal 1292
QY 1617 ATGGAGAAATAAAGCAAGAGAGAGGAAAGTGGATGGTTGAATAAAACTTTAAGGTG 1676
Db 1293 ThrGlyMetLeuAsnGluAlaGluGlyLysAlaIleLysLeuAlaLysAspValAlaSer 1312
QY 1677 GAGCAGGGAAGTCAATGGATGTGACGGGAAAGCTAATCGAGGAAAGCAAGAGCTTTTA 1736
Db 1313 LeuGlySerGlnLeuGlnAspThrGlnGluLeuGlnGluThrArgGlnLysLeu 1332
QY 1737 AAACCTCAATCT-----GAAATGGAGGAAAAGGAGTACAGTCTGCACAAAGGAGAGG 1787
Db 1333 AsnValSerThrLysLeuArgGlnLeuGluAspGluArgAsnSerLeuGlnAspGlnLeu 1352
QY 1788 GATGAGCTGATGGGTAAACTGAGGAGCGGAAGAAGAAAGGTCCTGTGAACCTGAGCTGCAGT 1847
Db 1353 AspGluGluMetGluAlaLysGlnAsnLeuGluArgHisValSerThrLeuAsnIleGln 1372
QY 1848 GTAGACTTACTAAAGAAGCGGCTT-----AATAAGATC-----GAT 1874
Db 1373 LeuSerAspSerLysLysLysLeuGlnAspPheAlaSerThrIleGluValMetGluGlu 1392
QY 1875 GGCATAGAGGAGGTAGAAAGGGAAATAAACCGAGGTAGTTCGTGCAAGGGTCTGAGTTC 1934
Db 1393 GlyLysLysArgLeuGlnLysGluMet---GluGlyLeuSerGlnGlnTyrGluGluLys 1411
QY 1935 ACCTGCCCGGAAGAC-----AATAAGATC----- 1958
Db 1412 AlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAspAsp 1431
QY 1959 -----AGAGAACTAACGCTTGAAATCGAGAGACTGAAGAAA 1994
Db 1432 LeuValValAspLeuAspAsnGlnArgGlnLeuValSerAsnLeuGluLysLysGlnLys 1451
QY 1995 CGGCTCCAGCAGTTGGAGGTGGTGAGGGGACTTG---ATGAAGACCGAGGACGAATAT 2051
Db 1452 LysPheAspGlnLeuLeuAlaGluGluLysAsnIleSerSerLysTyrAlaAspGluArg 1471
QY 2052 GACCAGTTGGAGCAGAAGTTCAGAACCGAGCAGGATAAGGCAAACTTCCTCTCCCAGCAG 2111
Db 1472 AspArgAlaGluAlaGluAlaArgGluLysGluThrLysAlaLeuSerLeuAlaArgAla 1491
QY 2112 CTCGAG-----GAAATCAAAACACCAAAATGGCCAAAGCACAAAGCCATAGAGAAAGGG 2162
Db 1492 LeuGluGluAlaLeuGluAlaLysGluGluLeuGluArgThrAsnLysMetLeuLysAla 1511
QY 2163 GAG-----GCCGTGAGCCAGGAAGCCGAACTGCGACACAGGTTTCGGCTGGAGGAG 2213
Db 1512 GluMetGluAspLeuValSerSerLysAspAspValGlyLysAsnValHisGluLeuGlu 1531
QY 2214 GCTAAAAGTCGTATTACAGGCCGAGGTGCAGGCTCTCAAGGAGAGAAGATCCACGAG--- 2270
Db 1532 LysSerLysArgAlaLeuGluThrGlnMetGluMetLysThrGlnLeuGluGluSer 1551
QY 2271 -----CTGATGAACAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTATTCGGTC 2321
Db 1552 GluAspAspValGlnAlaThrGluAspAlaLysLeuArgLeuGluValAsnMetGlnAla 1571
QY 2322 -----CTTCAGCAAAGATTTATGGAAGAAGAAACTAAGAAC 2357

Db 1572 LeuLysGlyGlnPheGluArgAspLeuGlnAlaArgAspGluGlnAsnGluGluLysArg 1591
QY 2358 AAGAACATGGGGAGGAGGTCTCTCAATCTGACCAAGAGCTAGAGCTTTCCAAGCGCTAC 2417
Db 1592 ArgGlnLeuGlnArgGlnLeuHisGluTyrGluThrGluLeuGlu---AspGluArgLys 1610
QY 2418 AGCCGAGCTCTC 2429
Db 1611 GlnArgAlaLeu 1614
RESULT 15
JC5420
smooth muscle myosin heavy chain 1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 02-Feb-2001
C;Accession: JC5420
R;Hasegawa, K.; Arakawa, E.; Oda, S.; Matsuda, Y.
Biochem. Biophys. Res. Commun. 232, 313-316, 1997
A;Title: Molecular cloning and expression of murine smooth muscle myosin heavy chains.
A;Reference number: JC5420; MUID:97242182; PMID:9125171
A;Accession: JC5420
A;Molecule type: mRNA
A;Residues: 1-1972 <HAS>
A;Cross-references: DDBJ:D85923; NID:G1945077; PIDN:BAA19690.1; PID:G1945078
A;Experimental source: smooth muscle
C;Comment: This protein plays a role in smooth muscle cell contraction.
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: nucleotide binding; P-loop
F;88-771/Domain: myosin motor domain homology <MMOT>
F;178-185/Region: nucleotide-binding motif A (P-loop)
Alignment Scores:
Pred. No.: 1.56e-13 Length: 1972
Score: 449.00 Matches: 203
Percent Similarity: 42.82% Conservative: 167
Best Local Similarity: 23.50% Mismatches: 292
Query Match: 5.81% Indels: 202
DB: 2 Gaps: 33
US-10-788-793-1 (1-4364) x JC5420 (1-1972)
QY 276 ACTAAGAAAGTCTGTGGAGTTATCCAAGGAGGACCTCATCCAG----- 317
Db 807 ThrLysArgGlnGlnGlnLeuThrAlaMetLysValIleGlnArgAsnCysAlaAlaTyr 826
QY 318 -----CTCCTGAGT 326
Db 827 LeuLysLeuArgAsnTrpGlnTrpArgLeuPheThrLysValLysProLeuLeuGln 846
QY 327 ATC-----ATGGAAGGGGAGTTGCAGGCTCGAGAAGATGTATCCACATGCTG----- 374
Db 847 ValThrArgGlnGluGluMetGlnAlaLysGluGluMetGlnLysIleThrGlu 866
QY 375 AGGACAGAGAAAAACCAAGCCCGAGGTTCTGGAGGCACACTATGGATCTGCAGAACCTGAG 434
Db 867 ArgGlnGlnLysAlaGluThrGluLeuLysGluLeuGlu----- 879
QY 435 AAAGTGCTTCGGGTCCTGCACCGAGATGCCATCCTTGCTCAAGAGAAGTCCATAGGAGAA 494
Db 880 -----GlnLysHisThrGlnLeuAlaGluGluLysThrLeu----- 891
QY 495 GAGTCTATGAGAAAACCTATCTCAGAGCTGGACAGACTGGAGGAAAAGCAGAGAGACG 554
Db 892 ---LeuGlnGluGlnLeuGlnAlaGluThrGluLeuTyrAlaGluSerGluGluMetArg 910
QY 555 TACCGCCGCATGCTAGAGCAGCTGCTGCTGGCTGAGAAGTGTCACAGCGCCACCGTGTCAC 614
Db 911 ValArg-----LeuAlaAlaLysLys-----Gln 918
QY 615 GAGCTGGAGAACGAGAAAGCACAAAGCACACTGACTACATGAACAAGAGCGACGACTTCACC 674
Db 919 GluLeuGluGluIleLeuHisGluMetGluAlaArgLeuGluGluGluAspArgArg 938

QY 675 AACCTGCTGGAGCAGGAGCGAGAGAGG-----TTGAAAAAGCTC 713
Db 939 GlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGlnMetLeuAspLeuGluGln 958
QY 714 CTTGAACAAGAAAAAGCTTACCAAGCCCGCAAGAAAGAAAGAAACGCT-----AAG 764
Db 959 LeuGluGluGluAlaAlaArgGlnLysLeuGlnLeuGluLysValThrAlaGluAla 978
QY 765 CGGCTCAACAACCTTCGAGATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCATGTTGGTG 824
Db 979 LysIleLysLysLeuGluAspAspIle-----LeuValMetAsp 991
QY 825 GACGAGAGGCAGATGCACATCGAGCAACTGGCCCTGCAGAGTCAGAAAGTCCAGGACCTC 884
Db 992 AspGlnAsnSerLysLeuSerLysGluArgLysLeuLeuGluGluArgValSerAspLeu 1011
QY 885 ACTCAGAAGCTGAGGGAGGAGGAAGAAAACTCAAAAGCGGTCACT----- 929
Db 1012 ThrThrAsnLeuAlaGluGluGluLysAlaLysAsnLeuThrLysLeuLysSerLys 1031
QY 930 -----TACAAATCCAAGAAAGACCGCCAG 953
Db 1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051
QY 954 AAGCTGCTCAAGTTAGAGTGGACTTCGAACACAAAGGCTCGAGTTTTCAGGAGCAC 1013
Db 1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071
QY 1014 GAAGAGATGAACGCCAAATTGGCGAATCAAGAATCTCACAAACCGCAACTTCGACTCAA 1073
Db 1072 AlaAspLeuGlnAlaGlnIleAla-----GluLeuLysMetGln 1084
QY 1074 CTG-----GTTGGCTTATCGCAAGGATTGAGGAGCTGGAA 1109
Db 1085 LeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgLeuAspGluGluIleAla 1104
QY 1110 GAGACCAATAAAAGCCTTCAGAAG-----GCAGAG 1139
Db 1105 GlnLysAsnAsnAlaLeuLysLysIleArgGluLeuGluGlyHisIleSerAspLeuGln 1124
QY 1140 GAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAACTCCAGTCTC 1199
Db 1125 GluAspLeuAspSerGluArgAlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeu 1144
QY 1200 ATGGCGGAAGTGGAGAGTCTGCGCAAGCGCTG-----CTTGAGATGGAGGCAAG 1250
Db 1145 GlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeuAspSerThrAlaThr 1164
QY 1251 GATGAAGAGATC---ACGAAGACCGAGGCCCCAGTGC CGGAGCTGAAGAAGAGCTCCAA 1307
Db 1165 GlnGlnGluLeuArgAlaLysArgGluGlnGluValThrValLeuLysLysAlaLeuAsp 1184
QY 1308 GAGGAA---GAACACCACAGCAAGGAACCTTAGACTAGAAGTGGAGAGCTGCAGAAAGAGG 1364
Db 1185 GluGluThrArgSerHisGluAlaGlnValGlnGluMetArgGlnLysHisThrGlnAla 1204
QY 1365 ATGCTCTGAGCTGGAGAAAGCTGGAGGAAGCGTTTCAGCCGAGTAAGTCGGAATGCACCCAG 1424
Db 1205 ValGluGluLeuThrGluGlnLeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLys 1224
QY 1425 CTCCATCTGAACCTGGAGAAGGAGAAGAACCTTAACCAAGACCTGCTGAACGAGCTGGAG 1484
Db 1225 SerLysGlnThrLeuGluLysGlu-----AsnAlaAspLeuAlaGlyGluLeuArg 1241
QY 1485 GTG-----GTCAAGAGTCGAGTTAAAGAACTCGAA----- 1514
Db 1242 ValLeuGlyGlnAlaLysGlnGluValGluHisLysLysLysLeuGluValGlnLeu 1261
QY 1515 -----TGCTCCGAGAGTAGACTGGAGAAGGCCGAGTTAAGCCTCAA 1556
Db 1262 GlnAspLeuGlnSerLysCysSerAspGlyGluArgAlaArgAlaGluLeuSer----- 1279
QY 1557 GATGACCTTACAAAGCTGAAAGTCTCTTCACTGTGATGCTGGTGGATGAGAGGAAAAATATG 1616

Db 1280 AspLysValHisLysLeuGln-----AsnGluValGluSerVal 1292
QY 1617 ATGGAGAAAAATAAAGCAAGAAAGAGAGGAAAGTGGATGGGTTGAATAAAAACTTTAAGGTG 1676
Db 1293 ThrGlyMetLeuAsnGluAlaGluGlyLysAlaIleLysLeuAlaLysAspValAlaSer 1312
QY 1677 GACGAGGAAAAAGTCATGGATGTGACGGAAAAAGCTAATCGAGGAAAGCAAGAGCTTTTA 1736
Db 1313 LeuGlySerGlnLeuGlnAspThrGlnGluLeuLeuGlnGluThrArgGlnLysLeu 1332
QY 1737 AAACCTCAAATCT-----GAAATGGAGGAAAGAGTACAGTCTGACAAAGGAGAGG 1787
Db 1333 AsnValSerThrLysLeuArgGlnLeuGluAspGluArgAsnSerLeuGlnAspGlnLeu 1352
QY 1788 GATGAGCTGATGGTAAACTGAGGAGCGAAGAAAGAGTCTGTGAACCTGAGCTGCAGT 1847
Db 1353 AspGluGluMetGluAlaLysGlnAsnLeuGluArgHisValSerThrLeuAsnIleGln 1372
QY 1848 GTAGACTTACTAAAGAACGGCTT-----AATAAGATC-----GAT 1874
Db 1373 LeuSerAspSerLysLysLysLeuGlnAspPheAlaSerThrIleGluValMetGluGlu 1392
QY 1875 GGCATAGAGGAGGTAGAAAGCGGAAATAAACCCAGGCTAGGTCTGCAAGGGGTCTGAGTTC 1934
Db 1393 GlyLysLysArgLeuGlnLysGluMet---GluGlyLeuSerGlnGlnTyrGluGluLys 1411
QY 1935 ACCTGCCCCGGAAGAC-----AATAAGATC----- 1958
Db 1412 AlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnLeuAspAsp 1431
QY 1959 -----AGAGAACTAAACGCTTTGAAATCGAGAGACTGAAGAAA 1994
Db 1432 LeuValValAspLeuAspAsnGlnArgGlnLeuValSerAsnLeuGluLysLysGlnLys 1451
QY 1995 CGGCTCCAGCAGTTGGAGGTGGTGGAGGGGACTTG---ATGAAGACCGAGGACGAATAT 2051
Db 1452 LysPheAspGlnLeuLeuAlaGluGluLysAsnIleSerSerLysTyrAlaAspGluArg 1471
QY 2052 GACCAGTTGGAGCAGAAGTTTCAGAACCCGAGCAGGATAAGGCAAACTTCCTCTCCAGCAG 2111
Db 1472 AspArgAlaGluAlaGluAlaArgGluLysGluThrLysAlaLeuSerLeuAlaArgAla 1491
QY 2112 CTCGAG-----GAAATCAAACACCAATGGCCCAAGCACAAAGCCCATAGAGAAAGGG 2162
Db 1492 LeuGluGluAlaLeuGluAlaLysGluGluLeuGluArgThrAsnLysMetLeuLysAla 1511
QY 2163 GAG-----GCCGTGAGCCAGGAAGCCGAACTGCGACACAGGTTTCGGCTGGAGGAG 2213
Db 1512 GluMetGluAspLeuValSerSerLysAspAspValGlyLysAsnValHisGluLeuGlu 1531
QY 2214 GCTAAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAGGAGAGATCCACGAG--- 2270
Db 1532 LysSerLysArgAlaLeuGluThrGlnMetGluGluMetLysThrGlnLeuGluGluSer 1551
QY 2271 -----CTGATGAACAAGGAAGACCCAGCTGTCTCAGCTCCAAGTCGACTATTCGGTC 2321
Db 1552 GluAspAspValGlnAlaThrGluAspAlaLysLeuArgLeuGluValAsnMetGlnAla 1571
QY 2322 -----CTTCAGCAAAAGATTTATGGAAGAGAACTAAGAAC 2357
Db 1572 LeuLysGlyGlnPheGluArgAspLeuGlnAlaArgAspGluGlnAsnGluGluLysArg 1591
QY 2358 AAGAACATGGGAGGAGGTCTCTCAATCTGACCAAGGAGCTAGAGCTTTCCAAAGCGCTAC 2417
Db 1592 ArgGlnLeuGlnArgGlnLeuHisGluTyrGluThrGluLeuGlu---AspGluArgLys 1610
QY 2418 AGCCGAGCTCTC 2429
Db 1611 GlnArgAlaLeu 1614

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
OM nucleic - protein search, using frame_plus_n2p model
Run on: September 7, 2004, 13:18:20 ; Search time 75.5 Seconds
(without alignments)
6019.448 Million cell updates/sec

Title: US-10-788-793-1
Perfect score: 7731
Sequence: 1 ccactgggtttctcaaggga.....aaaaaaaaaaaaaaaaaaaaa 4364

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10788793/runat_07092004 134020 10801/app query.fasta_1.4551
-DB=SwissProt 42 -QFMT=fastan -SUFFIX=resp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bites -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10788793 @CGN 1 1 118 @runat 07092004 134020 10801 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	476	6.2	2349	1 TPR_HUMAN	P12270 homo sapien
2	473.5	6.1	2663	1 CENE_HUMAN	Q02224 homo sapien
3	469	6.1	1960	1 MYH9_HUMAN	P35579 homo sapien
4	464.5	6.0	1790	1 USO1_YEAST	P25386 saccharomyc
5	458.5	5.9	1961	1 MYH9_RAT	Q62812 rattus norv
6	457	5.9	1938	1 MYS_AEQIR	P24733 aequipecten
7	452	5.8	1972	1 MYHB_HUMAN	P35749 homo sapien
8	449	5.8	1972	1 MYHB_MOUSE	O08638 mus musculu
9	447.5	5.8	1959	1 MYH9_CHICK	P14105 gallus gall
10	447	5.8	1427	1 REST_HUMAN	P30622 homo sapien
11	445.5	5.8	1407	1 TRHY_RABIT	P37709 oryctolagus
12	445	5.8	1972	1 MYHB_RABIT	P35748 oryctolagus
13	443	5.7	2017	1 MYSN_DROME	Q99323 drosophila
14	442.5	5.7	1433	1 REST_CHICK	O42184 gallus gall
15	441.5	5.7	1976	1 MYHA_BOVIN	Q27991 bos taurus
16	441	5.7	880	1 RA50_PYRAB	Q9uzc8 pyrococcus
17	441	5.7	1978	1 MYHB_CHICK	P10587 gallus gall
18	441	5.7	3210	1 CENF_HUMAN	P49454 homo sapien

19	440.5	5.7	1411	1 EEAL_HUMAN	Q15075 homo sapien
20	439	5.7	1940	1 MYH3_RAT	P12847 rattus norv
21	438	5.7	1976	1 MYHA_HUMAN	P35580 homo sapien
22	435.5	5.6	1976	1 MYHA_RAT	Q9jlt0 rattus norv
23	433.5	5.6	1102	1 MYSC_CHICK	P29616 gallus gall
24	433.5	5.6	1597	1 CTRO_MOUSE	P49025 mus musculu
25	433.5	5.6	1690	1 C190_DROME	Q9vj5es drosophila
26	433	5.6	1938	1 MYSS_CHICK	P13538 gallus gall
27	431.5	5.6	2230	1 GOA4_HUMAN	Q13439 homo sapien
28	431	5.6	2116	1 MYS2_DICDI	P08799 dictyosteli
29	430.5	5.6	1937	1 MYH8_HUMAN	P13535 homo sapien
30	430.5	5.6	2238	1 GOA4_MOUSE	Q91vw5 mus musculu
31	430	5.6	4473	1 PLE1_CRIGR	Q9ji55 cricetus
32	429	5.5	1898	1 TRHY_HUMAN	Q07283 homo sapien
33	429	5.5	3259	1 GOB1_HUMAN	Q14789 homo sapien
34	428.5	5.5	1940	1 MYH3_HUMAN	P11055 homo sapien
35	428	5.5	4687	1 PLE1_RAT	P30427 rattus norv
36	426.5	5.5	1938	1 MYHD_HUMAN	Q9ukx3 homo sapien
37	426	5.5	1940	1 MYH3_CHICK	P02565 gallus gall
38	426	5.5	3911	1 AKA9_HUMAN	Q99996 h a-kinase
39	425.5	5.5	2022	1 ANT1_ONCVO	P21249 onchocerca
40	425	5.5	1941	1 MYH2_HUMAN	Q9ukx2 homo sapien
41	420	5.4	978	1 RA50_AQUAE	O67124 aquifex aeo
42	417	5.4	879	1 RA50_SULTO	Q96yfr5 sulfolobus
43	417	5.4	1938	1 MYS_DCAEEL	P02567 caenorhabdi
44	415.5	5.4	1979	1 TRIA_HUMAN	Q15643 homo sapien
45	415	5.4	1360	1 CING_XENLA	Q9ptd7 xenopus lae

ALIGNMENTS

RESULT 1
TPR_HUMAN
ID TPR_HUMAN STANDARD; PRT; 2349 AA.
AC P12270;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Nucleoprotein TPR.
GN TPR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93064711; PubMed=1437155;
RA Mitchell P.J., Cooper C.S.;
RT "The human tpr gene encodes a protein of 2094 amino acids that has
RT extensive coiled-coil regions and an acidic C-terminal domain.";
RL Oncogene 7:2329-2333(1992).
RN [2]
RP REVISIONS, AND CHARACTERIZATION.
RX MEDLINE=95096166; PubMed=7798308;
RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
RA Sapphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
RT "Tpr, a large coiled coil protein whose amino terminus is involved in
RT activation of oncogenic kinases, is localized to the cytoplasmic
RT surface of the nuclear pore complex.";
RL J. Cell Biol. 127:1515-1526(1994).
RN [3]
RP SEQUENCE OF 1-142 FROM N.A.
RX MEDLINE=88262257; PubMed=3387099;
RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;
RT "tpr homologues activate met and raf.";
RL Oncogene 2:617-619(1988).
CC -!- FUNCTION: Component of the cytoplasmic fibrils of the nuclear pore
CC complex implicated in nuclear protein import. Its amino terminus
CC is involved in activation of oncogenic kinases.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE
CC COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH
CC TRP-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER
CC COMPONENTS, INCLUDING P62.

CC -!- TISSUE SPECIFICITY: Highest in testis, lung, thymus, spleen and
CC brain, lower levels in heart, liver and kidney.
CC -!- DISEASE: Involved in tumorigenic rearrangements with the MET, TRK
CC or RAF genes.
CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chromancer/Genes/TPRID282.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X66397; CAA47021.1; -.
DR EMBL; Y00672; CAA68681.1; -.
DR Genew; HGNC:12017; TPR.
DR MIM; 189940; -.
DR GO; GO:0005737; C:cytoplasm; TAS.
DR GO; GO:0005643; C:nuclear pore; TAS.
DR GO; GO:0006606; P:protein-nucleus import; TAS.
DR GO; GO:0006606; P:protein-nucleus import; TAS.
KW Coiled coil; Proto-oncogene; Chromosomal translocation;
KW Nuclear protein; Transport.
FT DOMAIN 78 360 COILED COIL (POTENTIAL).
FT DOMAIN 422 571 COILED COIL (POTENTIAL).
FT DOMAIN 575 628 COILED COIL (POTENTIAL).
FT DOMAIN 758 805 COILED COIL (POTENTIAL).
FT DOMAIN 834 869 COILED COIL (POTENTIAL).
FT DOMAIN 934 979 COILED COIL (POTENTIAL).
FT DOMAIN 1004 1064 COILED COIL (POTENTIAL).
FT DOMAIN 1138 1166 COILED COIL (POTENTIAL).
FT DOMAIN 1196 1241 COILED COIL (POTENTIAL).
FT DOMAIN 1262 1304 COILED COIL (POTENTIAL).
FT DOMAIN 1354 1434 COILED COIL (POTENTIAL).
FT DOMAIN 1476 1595 COILED COIL (POTENTIAL).
FT DOMAIN 527 530 POLY-SER.
FT DOMAIN 1833 1836 POLY-GLU.
FT DOMAIN 1957 1964 POLY-ASP.
FT DOMAIN 2295 2298 POLY-SER.
SQ SEQUENCE 2349 AA; 265600 MW; AFDD6885CEDCA9EF CRC64;

Alignment Scores:

Pred. No.: 1.57e-12 Length: 2349
Score: 476.00 Matches: 314
Percent Similarity: 36.85% Conservative: 252
Best Local Similarity: 20.44% Mismatches: 510
Query Match: 6.16% Indels: 460
DB: 1 Gaps: 67

US-10-788-793-1 (1-4364) x TPR_HUMAN (1-2349)

QY 174 GAAGATGCAAAAAGACAAAGGCCAATCGGAAGGAGGAG-----GATGTCATG 221
Db GluGluAlaArgGlyAsnHisValIleArgAspGluGluValSerSerAlaAspIleSer 527
QY 222 GCTTCCGGA---ACTATCAAAAGGCAC----- 245
Db SerSerSerGluValIleSerGlnHisLeuValSerTyrArgAsnIleGluGluLeuGln 547
QY 246 -----CTCAAAACCATCTGGAGAACTGAGAAAAGACT 278
Db GlnGlnAsnGlnArgLeuLeuValAlaLeuArgGluLeuGlyGluThrArgGluArgGlu 567
QY 279 AAGAAGTCTGTGGAGTTATCCAAG----- 302
Db GluGlnGluThrThrSerSerLysIleThrGluLeuGlnLeuLysLeuGluSerAlaLeu 587
QY 303 GAGGACCTCATCCAGTCCTGAGTATCATGGAAGGAGGAGTTGCAGGCTCGAGAAAGATGC 362
Db ThrGluLeuGluGlnLeuArgLysSerArgGlnHisGlnMetGlnLeuValAspSerIle 607
QY 363 ATC-----CACATGCTGAGGACAGAGAAAACCAAGCCCGAGGTTCTGGAG 407

Db ValArgGlnArgAspMetTyrArgIleLeuLeuSerGlnThrThrGlyValAlaIlePro 627
QY 408 GCACACTATGATCTGCAGAA----- 428
Db LeuHisAlaSerSerLeuAspAspValSerLeuAlaSerThrProLysArgProSerThr 647
QY 429 -----CCTGAGAAAAGTGCTT-----CGGGTCCTGCAC 455
Db SerGlnThrValSerThrProAlaProValProValIleGluSerThrGluAlaIleGlu 667
QY 456 CGAGATGCCATCCTTGCTCAAGAGAACTCCATAGGAGAACGCTCTATGAGAAACCTATC 515
Db AlaLysAlaAlaLeuLysGlnLeuGlnGluIlePheGluAsn--TyrLysLysGluLys 686
QY 516 TCAGAGCTGGACAGACTGGAGGAAAAGACAGAGGAGAGCTACCGC----- 560
Db AlaGluAsnGluLysIleGlnAsnGluGlnLeuGluLysLeuGlnGlnValThrAsp 706
QY 561 -----CGCATGCTAGAGCAGCTGCTGCTGGCTGAGAA----- 593
Db LeuArgSerGlnAsnThrLysIleSerThrGlnLeuAspPheAlaSerLysArgTyrGlu 726
QY 594 -----TGTCACAGCGCACCGCTGTACGAGCTGGAGAACGAGAAAG 632
Db MetLeuGlnAspAsnValGluGlyTyrArgArgGluIleThrSerLeuHisGluArgAsn 746
QY 633 CACAAGCACACTGACTACATGAACAAGAGCGAGCTTCACCAACCTGCTGGAGCAGGAG 692
Db GlnLysLeuThrAlaThrThrGlnLysGlnGluGlnIleIleAsnThrMetThrGlnAsp 766
QY 693 ---CGAGAGAGGTTGAAAAAGCTCCTTGAACAAGAA--AAAGCTTACCAAGCCCGCAAA 746
Db LeuArgGlyAlaAsnGluLysLeuAlaValAlaGluValArgAlaGluAsnLeuLysLys 786
QY 747 GAAAAGGAAAACGCTTAAG-----CGGTCAACAAACTTCGAGATGAGCTTGTG 794
Db GluLysGluMetLeuLysLeuSerGluValArgLeuSerGlnGlnArgGluSerLeu--- 805
QY 795 AAGCTCAAGTCTCTCGCCCTCATGTTGGTGGACGAGAGGCGAGATGCACATCGAGCAACTG 854
Db LeuAlaGluGlnArgGlyGlnAsnLeuLeuLeuThr 817
QY 855 GGCTGCGCAGAGT-----CAGAAAGTC 875
Db AsnLeuGlnThrIleGlnGlyIleLeuGluArgSerGluThrGluThrLysGlnArgLeu 837
QY 876 CAGGACCTCACTCAGAAAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 935
Db SerSerGlnIleGluLysLeuGluHisGluIleSerHisLeuLysLysLysLeuGluAsn 857
QY 936 TCCAAGGAAGACCGCCACGAAGCTGCTCAAG---TTAGAAGTGGAC---TTCGAACACAAG 989
Db GluValGluGlnArgHisThrLeuThrArgAsnLeuAspValGlnLeuLeuAspThrLys 877
QY 990 GCCTCGAGGTTTCCAGGAGGACCAAGAGATGAACGCCAAA----- 1031
Db ArgGlnLeuAspThrGluThrAsnLeuHisLeuAsnThrLysGluLeuLeuLysAsnAla 897
QY 1032 -----TTGGCGAATCAAGAATCTCACAAACCGCAACTTCGACTCAAACCTGGTTGGC 1082
Db GlnLysGluIleAlaThrLeuLysGlnHisLeuSerAsnMetGluValGlnValAlaSer 917
QY 1083 ---TTATCGCAAAGGATT-----GAGGAGCTGGAAGAGACC 1115
Db GlnSerSerGlnArgThrGlyLysGlyGlnProSerAsnLysGluAspValAspAspLeu 937
QY 1116 AATAAAGCCTTCAGAAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAATTCGCCAAA 1175
Db ValSerGlnLeuArgGlnThrGluGluGlnValAsnAspLeuLysGluArgLeuLysThr 957
QY 1176 GGGGAATGTGGAAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTCTT 1235

Db 958 SerThrSerAsnValGluGlnTyrGlnAlaMetValThrSerLeuGluGluSerLeuAsn 977

QY 1236 GAGATGGAGGGCAAGGATCAAGAGATCAGCAAGACC---GAGGCCCAGTGCCTGGAG--- 1289

Db 978 LysGluLysGlnValThrGluGluValArgLysAsnIleGluValArgLeuLysGluSer 997

QY 1290 -----CTGAAGAAGAAGCTCCAAGAGGAAGAACACCACAGCAAGGAA 1331

Db 998 AlaGluPheGlnThrGlnLeuGluLysLysLeuMetGluValGluLysGluLysGlnGlu 1017

QY 1332 CTT-----AGACTAGAAGTGGAGAAGCTGCAGAAAGAGGATGTCTGAGCTGGAG 1379

Db 1018 LeuGlnAspLysArgArgAlaIleGluSerMetGluGlnGlnLeuSerGluLeuLys 1037

QY 1380 AAG-----CTGGAGGAAGCGTTTCAGCCGG----- 1403

Db 1038 LysThrLeuSerSerValGlnAsnGluValGlnGluAlaLeuGlnArgAlaSerThrAla 1057

QY 1404 -----AGTAAGTCGGAATGC---ACCCAGCTCCATCTGAACCTGGAG 1442

Db 1058 LeuSerAsnGluGlnGlnAlaArgArgAspCysGlnGluAlaLysIleAlaValGlu 1077

QY 1443 AAGGAGAAGAACCTAACCAAGACCTGCTG-----AACGAGCTGGAG 1484

Db 1078 AlaGlnAsnLysTyrGluArgGluLeuMetLeuHisAlaAlaAspValGluAlaLeuGln 1097

QY 1485 GTGGTCAAGAGTCGAGTTAAAGAA----- 1508

Db 1098 AlaAlaLysGluGlnValSerLysMetAlaSerValArgGlnHisLeuGluThrThr 1117

QY 1509 -----CTCGAATGCTCCGAGAGTAGACTGGAGAAGGCCGAGTTA 1547

Db 1118 GlnLysAlaGluSerGlnLeuLeuGluCysLysAlaSer---TrpGluGluArgGluArg 1136

QY 1548 AGCCTCAAAGATGACCTTACAAAG-----CTGAAGTCCTTCACT 1586

Db 1137 MetLeuLysAspGluValSerLysCysValCysArgCysGluAspLeuGluLysGlnAsn 1156

QY 1587 GTGATGCTGGTGATGAGCGAANAATATGATGGAGAAAATAAGCAAGAGAGGAAA 1646

Db 1157 ArgLeuLeuHisAspGlnIleGluLysLeuSerAspLysValValAlaSerValLysGlu 1176

QY 1647 ---GTGGATGGG---TTGAATAAAAACTTTAAGTGGAG-----CAGGGAAAAGTC 1691

Db 1177 GlyValGlnGlyProLeuAsnValSerLeuSerLeuGluGluGlyLysSerGlnGluGlnIle 1196

QY 1692 ATGGATGTG-----ACGGAAAAGCTAATCGAGGAAGCAAG----- 1727

Db 1197 LeuGluIleLeuArgPheIleArgArgGluLysGluIleAlaGluThrArgPheGluVal 1216

QY 1728 -----AAGCTTTTAAAACTCAAATCTGAAATGGAG 1757

Db 1217 AlaGlnValGluSerLeuArgTyrArgGlnArgValGluLeuLeuGluArgGluLeuGln 1236

QY 1758 GAAAAGGAGTACAGTCTGACAAAGGAGAGG----- 1787

Db 1237 GluLeuGluAspSerLeuAsnAlaGluArgGluLysValGlnValThrAlaLysThrMet 1256

QY 1788 -----GATGAGCTGATGGGTAAA----- 1805

Db 1257 AlaGlnHisGluGluLeuMetLysLysThrGluThrMetAsnValValMetGluThrAsn 1276

QY 1806 -----CTGAGGAGCGAAGAAAGG----- 1826

Db 1277 LysMetLeuArgGluGluLysGluArgLeuGluGlnAspLeuGlnMetGlnAlaLys 1296

QY 1827 -----TCCTGTGAACCTGAGCTGC 1844

Db 1297 ValArgLysLeuGluLeuAspIleLeuProLeuGlnGluAlaAsnAlaGluLeuSerGlu 1316

QY 1845 AGTGTAGACTTACTA-----AAGAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGG 1895

Db 1317 LysSerGlyMetLeuGlnAlaGluLysLysLeuLeuGlu-----GluAspValLysArg 1334

QY 1896 GAAATAAACCGAGGTAGTTCGTGCAAGGGGTCTGAGTTACCTGCCCGAAGACAAATAAG 1955

Db 1335 TrpLysAlaArgAsnGlnHisLeuVal---SerGlnGlnLysAspProAspThrGluGlu 1353

QY 1956 ATCAGAGAACTAACGCTTGAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTG----- 2009

Db 1354 TyrArgLysLeuLeuSerGluLysGluValHisThrLysArgIleGlnGlnLeuThrGlu 1373

QY 2009 ----- 2009

Db 1374 GluIleGlyArgLeuLysAlaGluIleAlaArgSerAsnAlaSerLeuThrAsnAsnGln 1393

QY 2010 -----GAGGTGGTGGAGGGGACTTGTATGAAGACCGAGGACGAATATGAC----- 2054

Db 1394 AsnLeuIleGlnSerLeuLysGluAspLeuAsnLysValArgThrGluLysGluThrIle 1413

QY 2055 -----CAGTTGGAGCAGAAAGTTTCAGAACCGAGCAGGATAAAGGCAAACTTCTCTCCAG 2108

Db 1414 GlnLysAspLeuAspAlaLysIleIleAspIleGlnGluLysValLysThrIleThrGln 1433

QY 2109 -----CAGCTCGAGGAAATCAAAACACCAAAATGGCC 2138

Db 1434 ValLysLysIleGlyArgArgTyrLysThrGlnTyrGluGluLeuLysAlaGlnGlnAsp 1453

QY 2139 AAG-----CACAAAGCCATAGAGAAAGGGAG 2165

Db 1454 LysValMetGluThrSerAlaGlnSerSerGlyAspHisGln-----GluGlnHisVal 1471

QY 2166 GCCGTGAGCCAGGAAGCCGAACTGCGACACAGGTTTCGGCTGGAGGAGGTAAAAGTCGT 2225

Db 1472 SerValGlnGluMetGlnGluLeuLysGluThrLeuAsnGlnAlaGluThrLysSerLys 1491

QY 2226 GATTACAGGCCGAGGTGCAG-----GCTCTCAAGGAGAAGATCCACGAGCTG 2273

Db 1492 SerLeuGluSerGlnValGluAsnLeuGlnLysThrLeuSerGluLysGluThrGluAla 1511

QY 2274 ATGAACAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTATTTCGGTCTTCAGCAAGA 2333

Db 1512 ArgAsnLeuGlnGluGlnThrValGlnLeuGlnSerGluLeuSerArgLeuArgGlnAsp 1531

QY 2334 TTTATGGAAGAAACTAAGAACAAACATGTCGGAGGGAGGTCCTCAATCTGACCAAG 2393

Db 1532 LeuGlnAspArgThrThrGlnGluGluGlnLeuArgGlnGlnIle-----ThrGlu 1548

QY 2394 GAGCTAGAGCTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAAGG 2453

Db 1549 LysGluGluLysThrArgLys----- 1555

QY 2454 ATGGTGGACGTGCCTGTGGCCTCCACTGGGGTGAGACCGAGGCGGTGTGCGGGGATGCT 2513

Db 1556 -----AlaIleValAlaLysSerLysIleAlaHisLeuAlaGlyValLysAspGln 1573

QY 2514 GCGGAGGAGGAGACCCCGGCTGTGTTTCATTTCGCAAAATCCTTCCAGGAGGAAAATCACATC 2573

Db 1574 LeuThrLysGluAsnGluGluLeuLysGlnArgAsnGlyAlaLeuAspGlnGlnLysAsp 1593

QY 2574 ATGAGTAATCTTCACAGGTAGGCCCTGAAGAAACCCCATGGAA---CGGTCTCGGTCCTC 2630

Db 1594 GluLeuAspValArgIleThrAlaLeuLysSerGlnTyrGluGlyArgIleSerArgLeu 1613

QY 2631 GACAGGTATCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTGATTCTCTGGATG 2690

Db 1614 GluArg-----GluLeuArgGluHisGlnGluArgHisLeu 1625

QY 2691 AGAAAAAGAGAAAACGGTCTTCCACTCCGAGGAGAAAGGCCAGGCCAACAGGGGT 2750

Db 1626 GluGlnArgAspGlu---ProGlnGluProSerAsnLys----- 1637

QY 2751 GCAGGGCACCCCGGGAGCTGGTCTTAGCACCAAGAGCCAGGCCCTACACATCCGT 2810

Db 1638 -----ValProGluGlnGlnArgGlnIleThrLeuLys 1648


```
DR PROSITE; PS50067; KINESIN MOTOR DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
  Cell cycle; Centromere; Lipoprotein; Prenylation.
FT DOMAIN 1 335 KINESIN-MOTOR.
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP_BIND 86 93 ATP (BY SIMILARITY).
FT LIPID 2660 2660 S-farnesyl cysteine.
SQ SEQUENCE 2663 AA; 312087 MW; CEFCL380C8C8CB8 CRC64;

Alignment Scores:
Pred. No.: 1.98e-12 Length: 2663
Score: 473.50 Matches: 289
Percent Similarity: 39.37% Conservative: 263
Best Local Similarity: 20.61% Mismatches: 461
Query Match: 6.12% Indels: 389
DB: 1 Gaps: 64

US-10-788-793-1 (1-4364) x CENE_HUMAN (1-2663)
QY 171 TCAGAAGATGCAAAAAGAAACAAGGCCAATCGGAAGGAGGAGGAT----- 215
  |||::: ::::: ||| |||||:::|
Db 1382 SerGlnSerLysGlnGluGlnSerLeuAsnMetLysGluLysAspAsnGluThrThrLys 1401
QY 216 -----GTCATGGCTTCGGAACTATCAAAAGGCACCTC 248
  ::| ||| |||::: ||| ::::: |||
Db 1402 IleValSerGluMetGluGlnPheLysProLysAspSerAlaLeuLeuArgIleGluIle 1421
QY 249 AAACCATCTGGAGAAAGTGAGAAAAAGACTAAG-----AAGTCTGTGGAG 293
  ::| ||| |||::: ||| ::::: |||
Db 1422 GluMetLeuGlyLeuSerLysArgLeuGlnGluSerHisAspGluMetLysSerValAla 1441
QY 294 TTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGGAGTTG---CAGGCT 350
  |||:::| |||:::| |||:::| |||:::| |||:::| |||
Db 1442 LysGluLysAspAspLeuGlnArgLeuGlnGluValLeuGlnSerGluSerAspGlnLeu 1461
QY 351 CGAGAAGATGTCATCCACATGCTGAGGACAGAG---AAACCAAGCCCGAGGTTCTCGAG 407
  ::| |||::: |||::: |||::: |||::: |||::: |||::: |||
Db 1462 LysGluAsnIleLysGluIleValAlaLysHisLeuGluThrGluGluLeuLysVal 1481
QY 408 GCACACTATGGATCTGCAGAACTTGAGAAAGTGTCTCGGTCCTGCACCGAGATGCCATC 467
  ||||| ||| |||::: |||::: |||::: |||::: |||::: |||
Db 1482 AlaHisCysCysLeuLysGluGlnGluGluThrIleAsnGluLeuArg-----Val 1498
QY 468 CTGTGCTCAAGAGAAGTCCATAGGAGAAGACGCTCTATGAGAAACCTATCTCAGAGCTG--- 524
  ||||| ||||| ::::: |||::: |||::: |||::: |||
Db 1499 AsnLeuSerGluLysGluThrGluIleSerThrIleGlnLysGlnLeuGluAlaIleAsn 1518
QY 525 GACAGACTGGAGGAAAAGCAGAAAGGAGACGTACCCGCCGCTAGAGCAGCTGTGCTG 584
  |||:::| |||:::| |||:::| |||:::| |||:::| |||:::| |||
Db 1519 AspLysLeuGlnAsnLysIleGlnGluIleTyrGluLysGlu---GluGlnLeuAsnIle 1537
QY 585 GCTGAG-----AAGTGTACAGG 602
  ::| ||| |||||
Db 1538 LysGlnIleSerGluValGlnGluAsnValAsnGluLeuLysGlnPheLysGluHisArg 1557
QY 603 CGC-----ACCGTGTACGAGCTGGAGAACGAGAAGCACAAAGCACACTGACTAC 650
  ::| ::::: |||::: |||::: |||::: |||::: |||
Db 1558 LysAlaLysAspSerAlaLeuGlnSerIleGluSerLysMetLeuGluLeuThrAsnArg 1577
QY 651 ATGAACAAGAGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGAGAGGTTGAAAAAG 710
  ::| :::| ||| ::::: |||::: |||::: |||::: |||::: |||
Db 1578 LeuGlnGluSerGlnGluGluIleGlnIleMetIleLysGluLysGluMetLysArg 1597
QY 711 CTCCTTGAAACAAGAAAAGCTTACCAAGCCCGCAAGAA-----AAGGAAAACGCTAAG 764
  ::| :::| ||| ||| ::::: |||::: |||::: |||
Db 1598 Val-----GlnGluAlaLeuGlnIleGluArgAspGlnLeuLysGluAsnThrLys 1614
QY 765 CGGCTC---AACAAACTTCAGATGAGCTTGTGAAG-----CTCAAGTCCTTC 809
  ::| |||:::| |||::: |||
Db 1615 GluIleValAlaLysMetLysGluSerGlnGluLysGluTyrGlnPheLeuLysMetThr 1634
QY 810 GCCCTCATGTTGGTGGACGAGAGGCAGATGCACATCGAGCAACTG-----GGCCTGCAG 863
```

```

  |||::: |||::: ||||| ||||| |||
Db 1635 AlaValAsnGluThrGlnGluLysMetCysGluIleGluHisLeuLysGluGlnPheGlu 1654
QY 864 AGTCAGAAAAGTCCAG-----GACCTCACTCAGAAGCTG 896
  ::| |||||::: ||||| ||||| |||
Db 1655 ThrGlnLysLeuAsnLeuGluAsnIleGluThrGluAsnIleArgLeuThrGlnIleLeu 1674
QY 897 AGGAGAGGAGAAACTCAAAGCGGTCACTTACAAATCCAAGGAAGACCGC----- 950
  ||| |||:::| |||:::| |||:::| |||:::| |||:::| |||
Db 1675 HisGluAsnLeuGluMetArgSerValThr---LysGluArgAspAspLeuArgSer 1693
QY 951 ---CAGAAGCTGCTCAAGTTAGAAGTGGAC-----TTCGAACACAAAGGCC 992
  ::::: |||||:::| |||
Db 1694 ValGluGluThrLeuLysValGluArgAspGlnLeuLysGluAsnLeuArgGluThrIle 1713
QY 993 TCGAGGTTTTCCAGGAGCACGAAGAGATGAACGCCAAATTTGGCAATCAAGAAATCTCAC 1052
  ::| |||::: |||::: |||::: |||::: |||::: |||::: |||
Db 1714 ThrArgAspLeuGluLysGlnGluGluLeuLysIleValHisMetHisLeuLysGluHis 1733
QY 1053 AACCGGCAACTTCGACTCAAACTGGTTGGCTTA---TCGCAAAAGGATTGAGGAGCTGGAA 1109
  ::| ||||| ||||| ||||| ||||| |||||
Db 1734 GlnGluThrIleAsp---LysLeuArgGlyIleValSerGluLysThrAsnGluIleSer 1752
QY 1110 GAGACCAATAAAAGCCTTCAGAAAGGCAGAGGAAAGAGCTCCAG-----GAGCTG 1157
  ||| |||||::: |||::: |||::: |||::: |||::: |||::: |||
Db 1753 AsnMetGlnLysAspLeuGluHisSerAsnAspAlaLeuLysAlaGlnAspLeuLysIle 1772
QY 1158 AGAGAGAAAATTGCCAAAGGGGAATGTGGAAACTCTCAGTCTCATGGCGGAAGTGGAGACT 1217
  ::| |||::: |||::: |||::: |||::: |||::: |||::: |||
Db 1773 GlnGluGluLeuArgIleAlaHisMetHisLeuLysGluGlnGlnGluThrIleAspLys 1792
QY 1218 CTGCGCAAGCGCGTGTGAGATGGAGGGCAAG-----GATGAAGAGATCACG 1265
  ||||| ||||| ||||| ||||| |||||
Db 1793 LeuArgGlyIleValSerGluLysThrAspLysLeuSerAsnMetGlnLysAspLeuGlu 1812
QY 1266 AAGACCGAGGCCAGTCGCGGAGCTGAAGAAGAGTCCAAAGAGGAAGAACACACACAGC 1325
  ::| |||||::: |||::: |||::: |||::: |||::: |||::: |||
Db 1813 AsnSerAsnAlaLysLeuGlnGluLysIleGlnGluLeuLysAlaAsnGluHisGlnLeu 1832
QY 1326 AAGAACTTAGACTAGAAAGTGGAGAAGCTGCAGAAAGAGGATGTCTGAGCTGGAGAAGCTG 1385
  |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
Db 1833 IleThrLeuLysLysAspValAsnGluThrGlnLysLysValSerGluMetGluGlnLeu 1852
QY 1386 -----GAGAAAGCGTTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCAT 1430
  ::::: |||::: |||::: |||::: |||::: |||
Db 1853 LysLysGlnIleLysAspGlnSerLeuThrLeuSerLysLeuGluIleGluAsnLeu--- 1871
QY 1431 CTGAACCTGGAGAGGAGAAGAACCTTAACCAAGCTGCTGAACGAGCTGGAGGTGGTC 1490
  ||||| ::::: |||::: |||::: |||::: |||::: |||::: |||
Db 1872 -----AsnLeuAlaGlnGluLeuHisGluAsnLeuGluGluMet 1884
QY 1491 AAGAGTCGAGTTAAAGAA-----CTCGAATGCTCCGAGAGTAGACTGGAGAAGGCC 1541
  ||||| ::||| ||||| ||||| |||||
Db 1885 LysSerValMetLysGluArgAspAsnLeuArgArgValGluGluThrLeuLysLeuGlu 1904
QY 1542 GAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAGCTCCTTCACTGTGATGCTGGTGGAT 1601
  |||||::: |||::: |||::: |||::: |||::: |||::: |||
Db 1905 ArgAspGlnLeuLysGluSerLeuGlnGluThrLysAlaArgAspLeuGluIleGlnGln 1924
QY 1602 GAGAGGAAAAATATGATGGAGAAAAATAAAGCAAGAAGAGAGGAAAGTGGATGGGTGAAT 1661
  ||| ||| |||||::: |||::: |||::: |||::: |||
Db 1925 GluLeuLysThrAlaArgMetLeuSerLysGluHisLysGluThrValAspLysLeuArg 1944
QY 1662 AAAAAGCTTTAAGGTGAGCAGGGGAAAAAGTCAATGATGTGACGGAAAAAGCTAATCGAGGAA 1721
  ::| |||::: |||::: |||::: |||::: |||::: |||
Db 1945 Glu-----LysIleSerGluLysThrIleGlnIleSerAspIle 1957
QY 1722 AGCAAGAAGCTTTAAAACTCAAATCTGAAATGGAGGAAAAAGGAGTACAGTCTGACAAAAG 1781
  ||| ||| ||| |||||::: |||::: |||::: |||::: |||
Db 1958 GlnLysAspLeuAspLysSerLysAspGluLeuGlnLysLysIleGlnGluLeuGlnLys 1977
QY 1782 GAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGCAAGAAAGGTCTCTGTAAGTCTGAGC 1841
  ::| ::| ||| ||||| ||||| |||||
```

Db 1978 LysGluLeuGlnLeu-----LeuArgValLysGluAsp----- 1988
QY 1842 TGCAGTGTAGACTTACTAAAGACGGCTTGATGGCATAGAGGAGGTAGAAAGGAAATA 1901
Db 1989 -----ValAsnMetSerHisLysLysIleAsnGluMetGluGlnLeuLysLysGlnPhe 2006
QY 1902 AACCGAGGTAGGTGCTGCAAGGGGTCTGAGTTACCTGTCCCGGAAGACAATAAGATCAGA 1961
Db 2007 GluProAsnTyrLeuCysLys-----CysGluMetAspAsn-----Phe 2019
QY 1962 GAACTAACGCTTGAAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTGGTGAG 2021
Db 2020 GlnLeuThr-----LysLysLeuHisGluSerLeuGluGluIleArgIleVal--- 2035
QY 2022 GGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAGCAGAAGTTCAGAACCGAG 2081
Db 2036 -----AlaLysGluArgAspGluLeuArgArgIleLysGluSerLeuLysMetGlu 2052
QY 2082 CAGGATAAGGCAAACTTCCTCTCCAGCAGCTCGAGGAATC-----AAA 2126
Db 2053 ArgAspGln-----PheIleAlaThrLeuArgGluMetIleAlaArgAspArgGlnAsn 2070
QY 2127 CACCAAAATGGCCAAGCACAAAAGCCATA----- 2153
Db 2071 HisGlnValLysProGluLysArgLeuLeuSerAspGlyGlnGlnHisLeuMetGluSer 2090
QY 2154 -----GAGAAA-----GGGAGGCCGTGAGCCAGGAAGCCGAACTGCGACAC 2195
Db 2091 LeuArgGluLysCysSerArgIleLysGluLeuLeuLysArgTyrSerGluMetAspAsp 2110
QY 2196 AGGTTTCGGCTGGAGGAGGTAAAGCTCGTGATTTACAGGCCGAGGTGCAG----- 2246
Db 2111 HisTyrGluCysLeuAsnArgLeuSerLeuAspLeuGluLysGluIleGluPheHisArg 2130
QY 2247 -----GCTCTCAAGGAGAAAGATCCAC 2267
Db 2131 IleMetLysLysLeuLysTyrValLeuSerTyrValThrLysIleLysGluGluGlnHis 2150
QY 2268 GAGCTGATGAACAAGGAAGACACGACTGTCTCAGCTCCAAGTCGACTATTCGGTCTTCAG 2327
Db 2151 GluCysIleAsnLys-----PheGlu 2157
QY 2328 CAAAGATTATGGAAGAAGAAAACTAAGAACAAAGAACATGGGGAGGGAGGTCTCTCAATCTG 2387
Db 2158 MetAspPheIleAspGluValGluLysGlnLysGluLeuLeuIleLysIleGlnHisLeu 2177
QY 2388 ACCAAGAGCTAGAGCTTTCCAAGCGGTACAGCCGAGTCTCAGGCCGAGTGGGAACGGC 2447
Db 2178 GlnGlnAspCysAspValProSerArgGluLeuArgAspLeuLysLeuAsnGlnAsn--- 2196
QY 2448 CGAAGGATGGTGACGTGCCTGTGGCCTCCACTGGGGTGCAGACCGGCGGTGTGCGGG 2507
Db 2197 -----MetAspLeuHisIle-----GluGluIleLeuLys 2206
QY 2508 GATGTCGGGAGGAGGAGACCCCGGCTGTGTTCATTGCAAAATCCTTCCAGGAGGAAAAAT 2567
Db 2207 AspPheSerGluSerGluPheProSer-----IleLysThrGluPheGlnGln----- 2222
QY 2568 CACATCATGAGTAATCTTCGACAGGTA----- 2594
Db 2223 ---ValLeuSerAsnArgLysGluMetThrGlnPheLeuGluGluTrpLeuAsnThrArg 2241
QY 2595 -----GGCCTGAAGAAACCCATGGAACGTCCTCGTCCCTC 2630
Db 2242 PheAspIleGluLysLeuLysAsnGlyIleGlnLysGluAsnAspArgIleCysGlnVal 2261
QY 2631 -----GACAGGTATCCCCCAGCAGCGAATGAG----- 2657
Db 2262 AsnAsnPhePheAsnAsnArgIleIleAlaIleMetAsnGluSerThrGluPheGluGlu 2281
QY 2658 -----CTCACCATGAGGAAGTCTTGGATTCTTGGATGAGAAAAAGAGAAAACGGTCCT 2711
Db 2282 ArgSerAlaThrIleSerLysGluTrp-----GluGlnAspLeu 2294

QY 2712 TCCACTCCGCAGAGAGAAAGGGCCCGAGCCCAAAACCAGGGTGCAGGGCACCCCGGGAGCTG 2771
Db 2295 LysSerLeuLysGluLysAsnGluLys----- 2303
QY 2772 GTCCTAGCACCAAGACGAGGCCAGCCCTACACATCCGTGTGACACCAGATCATGAGAAC 2831
Db 2304 -----LeupheLysAsnTyrGlnThrLeu-----LysThr 2313
QY 2832 AGCACTGCCACCCCTGGAGATCACAAAGCCCCACATCTGAAGAGTTTTTCTCTAGTACCACC 2891
Db 2314 SerLeuAlaSerGlyAlaGlnValAsnProThrThrGlnAsp----- 2327
QY 2892 GTCATTCTCTACCTTAGGCAACCCAGAAACCAAGAATAACCATATTATCCATCACCCCAATGTC 2951
Db 2328 -----AsnLysAsnProHisValThr---SerArgAlaThrGlnLeu 2340
QY 2952 ATGTCGCAAAAGCCCAAAAGTGCAGATCCTACTCTC-----GGCCCCAGAACGAGCCATG 3005
Db 2341 ThrThrGluLysIleArgGluLeuGluAsnSerLeuHisGluAlaLysGluSerAlaMet 2360
QY 3006 TCC-----CCTGTCCAGATTACTACTATT 3029
Db 2361 HisLysGluSerLysIleIleLysMetGlnLysGluLeuGluValThrAsnAspIleIle 2380
QY 3030 TCCAGAGAGAAGAGCCCGGAAGGTGGAAGGAGCGCTTTGCCAGACGCGCTGCATCCCCC 3089
Db 2381 AlaLysLeuGlnAlaLysValHisGluSerAsnLysCysLeuGluLysThrLysGluThr 2400
QY 3090 ATCCAATCATGACGGTGTCAACATCTGCAGCTCCCACTGAAATCGTGTCTCTCTCTGAA 3149
Db 2401 IleGlnValLeuGlnAspLysValAla-----LeuGlyAlaLysProTyr 2415
QY 3150 TCTCAGGAAGTG---CCTATGGGAAGGACTATCTCAAAGTCAACCCGGAACCAAACT 3206
Db 2416 LysGluGluIleGluAspLeuLysMetLysLeuValLysIleAspLeuGluLysMetLys 2435
QY 3207 GTTCCAGCCCCGTGCGGAAGTACRACTCCAATGCTAATATCATCACCACGGAAGACAAT 3266
Db 2436 AsnAlaLysGluPheGluLys---GluIleSerAlaThrLysAlaThrValGluTyrGln 2454
QY 3267 AAAATTCACATTCACCTG-----GGTTCTCAGTTTAAAGCGATCTCCT 3308
Db 2455 LysGluValIleArgLeuLeuArgGluAsnLeuArgArgSerGlnGlnAlaGlnAspThr 2474
QY 3309 GGGCTGCGGTGAAGCGGTGAGCCAGTTATACCGTCCGGCCTGTC----- 3356
Db 2475 SerValIleSerGluHisThrAspProGlnProSerAsnLysProLeuThrCysGlyGly 2494
QY 3357 -----AACGTGACAGCG 3368
Db 2495 GlySerGlyIleValGlnAsnThrLysAlaLeuIleLeuLysSerGluHisIleArgLeu 2514
QY 3369 GAGAAGGAGGTTTCT-----ACAGGCACAGTC 3395
Db 2515 GluLysGluIleSerLysLeuLysGlnGlnAsnGluGlnLeuIleLysGlnLysAsnGlu 2534
QY 3396 CTTCGCTCTCCCAGGAACCACTCTCTTCAAGA----- 3428
Db 2535 LeuLeuSerAsnAsnGlnHisLeuSerAsnGluValLysThrTrpLysGluArgThrLeu 2554
QY 3429 -----CCCGGTGCTAGCAAAAGTGACC 3449
Db 2555 LysArgGluAlaHisLysGlnValThrCysGluAsnSerProLysSerProLysValThr 2574
QY 3450 AGCACTATAACT-----ATAACCCCGGTCACAA-----CGTCAT----- 3483
Db 2575 GlyThrAlaSerLysLysLysGlnIleThrPro-SerGlnCysLysGluArgAsnLeuG1 2594
QY 3484 -----CCACACGAGGAACCCAATCAGTGT---CAGGACAAGATGGGTCTATCTCAG 3530
Db 2594 nAspProValProLysGluSerProLysSerCysPheAspSerArgSerLys-SerL 2614

QY 3531 CGGCCTACCCCGCCGTCATCTCTATGTCACAAAGGTATGAA----- 3571
|||||
Db 2614 euProSerProHisProValArgTyrPheAspAsnSerSerLeuGlyLeuCysProGluV 2634
QY 3572 -----AGCTGGAAAGCCAGTAGTG 3590
|||||
Db 2634 alGlnAsnAlaGlyAlaGluSerValAspSerGlnProGlyProTrpHisAlaSerSerg 2654
QY 3591 GC 3592
||
Db 2654 ly 2654
RESULT 3
MYH9_HUMAN
ID MYH9_HUMAN STANDARD; PRT; 1960 AA.
AC P35579; O60805;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,
type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
GN MYH9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,
Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
Mcclay J., McLaren S., Mcmurray A.A., Milne S.A., Mortimore B.J.,
Odeh C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
Fulton R., Johnson D., Benis G., Bentley D., Bradshaw H., Bourne S.,
Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
Scheet P., Walker C., Wamaley A., Wohldmann P., Pepin K., Nelson J.,
Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
Edelmann L., Kim U.-J., Shizuya E., Simon M.I., Dumanski J.P.,
Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
Khan A.S., Lane L., Tilahun Y., Wright H.,
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RN [2]

RP SEQUENCE OF 1-1337 FROM N.A.
RX MEDLINE=92003925; PubMed=1912569;
RA Toothaker L.E., Gonzalez D.A., Tung N., Lemons R.S., le Beau M.M.,
RA Arnaut M.A., Clayton L.K., Tenen D.G.;
RT "Cellular myosin heavy chain in human leukocytes: isolation of 5'
cDNA clones, characterization of the protein, chromosomal
localization, and upregulation during myeloid differentiation.";
RL Blood 78:1826-1833(1991).
RN [3]
RP SEQUENCE OF 1-715 FROM N.A.
RX MEDLINE=91316803; PubMed=1860190;
RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,
RA Gdula D., Adelstein R.S., Weir L.;
RT "Human nonmuscle myosin heavy chains are encoded by two genes located
on different chromosomes.";
RL Circ. Res. 69:530-539(1991).
RN [4]
RP SEQUENCE OF 714-1960 FROM N.A.
RX MEDLINE=90138958; PubMed=1967836;
RA Saez C.G., Myers J.C., Shows T.B., Leinwand L.A.;
RT "Human nonmuscle myosin heavy chain mRNA: generation of diversity
through alternative polyadenylation.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:1164-1168(1990).
RN [5]
RP VARIANT DFNA17 HIS-705.
RX MEDLINE=20489856; PubMed=11023810;
RA Lalwani A.K., Goldstein J.A., Kelley M.J., Luxford W., Castelein C.M.,
RA Mhatre A.N.;
RT "Human nonsyndromic hereditary deafness DFNA17 is due to a mutation in
nonmuscle myosin MYH9.";
RL Am. J. Hum. Genet. 67:1121-1128(2000).
RN [6]
RP VARIANTS MHA/FTNS/SBS LYS-93; CYS-702; CYS-1165; HIS-1424 AND
LYS-1841.
RX MEDLINE=20428192; PubMed=10973259;
RA Seri M., Cusano M., Gangarossa S., Caridi G., Bordo D., Lo Nigro C.,
Ghiggeri G.M., Ravazzolo R., Savino M., Del Vecchio M., d'Apolito M.,
Iolascon A., Zelante L.L., Savoia A., Balduini C.L., Noris P.,
Magrini U., Belletti S., Heath K.E., Babcock M., Glucksman M.J.,
Aliprandis E., Bizzaro N., Desnick R.J., Martignetti J.A.;
RT "Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and
Sebastian syndromes.";
RL Nat. Genet. 26:103-105(2000).
RN [7]
RP VARIANTS MHA ILE-1155 AND LYS-1841.
RX MEDLINE=20428193; PubMed=10973260;
RA Kelley M.J., Jawien W., Ortel T.L., Korczak J.F.;
RT "Mutation of MYH9, encoding non-muscle myosin heavy chain A, in
May-Hegglin anomaly.";
RL Nat. Genet. 26:106-108(2000).
CC -!- FUNCTION: Cellular myosin appears to play a role in cytokinesis,
cell shape, and specialized functions such as secretion and
capping.
CC -!- SUBUNIT: Myosin is a hexameric protein that consists of 2 heavy
chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2
regulatory light chain subunits (MLC-2).
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
characteristic for alpha-helical coiled coils.
CC -!- DISEASE: Defects in MYH9 are the cause of May-Hegglin anomaly
(MHA) [MIM:155100]. MHA is an autosomal dominant
macrothrombocytopenia characterized by thrombocytopenia, giant
platelets and leukocyte inclusions.
CC -!- DISEASE: Defects in MYH9 are the cause of Fechtner syndrome (FTNS)
[MIM:153640]. FTNS is an autosomal dominant macrothrombocytopenia
characterized by thrombocytopenia, giant platelets and leukocyte
inclusions. With additional alport-like clinical features of
sensorineural deafness, cataracts and nephritis.
CC -!- DISEASE: Defects in MYH9 are the cause of Sebastian syndrome (SBS)
[MIM:605249]. SBS is an autosomal dominant macrothrombocytopenia
characterized by thrombocytopenia, giant platelets and leukocyte
inclusions.
CC -!- DISEASE: Defects in MYH9 are the cause of autosomal dominant

CC nonsyndromic sensorineural deafness 17 (DFNA17) [MIM:603622].
CC DFNA17 is characterized by progressive hearing impairment and
CC cochleosaccular degeneration.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; Z82215; -; NOT ANNOTATED_CDS.
CC DR EMBL; M81105; AAA59888.1; -.
CC DR EMBL; M69180; AAAG1765.1; -.
CC DR EMBL; M31013; AAA36349.1; -.
CC DR PIR; A61231; A61231.
CC DR HSSP; P10587; 1BR2.
CC DR Genew; HGNC:7579; MYH9.
CC DR MIM; 160775; -.
CC DR MIM; 155100; -.
CC DR MIM; 153640; -.
CC DR MIM; 605249; -.
CC DR MIM; 603622; -.
CC DR GO; GO:0005860; C:non-muscle myosin; TAS.
CC DR InterPro; IPR000048; IQ_region.
CC DR InterPro; IPR001609; myosin_head.
CC DR InterPro; IPR004009; Myosin_N.
CC DR InterPro; IPR002928; Myosin_tail.
CC DR InterPro; IPR002017; Spectrin.
CC DR Pfam; PF00612; IQ; 1.
CC DR Pfam; PF00063; myosin_head; 1.
CC DR Pfam; PF02736; Myosin_N; 1.
CC DR Pfam; PF01576; Myosin_tail; 1.
CC DR PRINTS; PR00193; MYOSINHEAVY.
CC DR ProDom; PD000355; myosin_head; 1.
CC DR SMART; SM00015; IQ; 1.
CC DR SMART; SM00242; MYSC; 1.
CC DR PROSITE; PS50096; IQ; 1.
CC DR Myosin; ATP-binding; Calmodulin-binding; Actin-binding; Coiled coil;
CC Multigene family; Disease mutation; Deafness.
CC FT DOMAIN 1 778 MYOSIN HEAD-LIKE.
FT DOMAIN 779 808 IQ.
FT DOMAIN 837 1926 COILED COIL (POTENTIAL).
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 654 676 ACTIN-BINDING.
FT VARIANT 93 93 N -> K (IN MHA).
FT /FTID=VAR 010791.
FT VARIANT 702 702 R -> C (IN FTNS).
FT /FTID=VAR 010792.
FT VARIANT 705 705 R -> H (IN DFNA17).
FT /FTID=VAR 010793.
FT VARIANT 1155 1155 T -> I (IN MHA).
FT /FTID=VAR 010794.
FT VARIANT 1165 1165 R -> C (IN SBS).
FT /FTID=VAR 010795.
FT VARIANT 1424 1424 D -> H (IN FTNS).
FT /FTID=VAR 010796.
FT VARIANT 1841 1841 E -> K (IN MHA).
FT /FTID=VAR 010797.
FT CONFLICT 53 55 EAI -> RGH (IN REF. 3).
FT CONFLICT 660 660 T -> S (IN REF. 3).
FT CONFLICT 869 869 T -> M (IN REF. 4).
FT CONFLICT 931 931 C -> Y (IN REF. 4).
FT CONFLICT 1240 1241 KG -> GR (IN REF. 4).
FT CONFLICT 1350 1350 E -> EE (IN REF. 2).

Alignment Scores:
Pred. No.: 3e-12 Length: 1960
Score: 469.00 Matches: 220
Percent Similarity: 41.88% Conservative: 177

Best Local Similarity: 23.21% Mismatches: 327
Query Match: 6.07% Indels: 224
DB: 1 Gaps: 37
US-10-788-793-1 (1-4364) x MYH9_HUMAN (1-1960)
QY 459 GATGCCATCCTTGCTCAAGAGAAAGTCCATAGGAGAGACGCTATGAGAAACCTATCTCA 518
Dd 845 GluGluMetAlaLysGluGluGluLeuValLys---ValArgGluLysGlnLeuAla 863
QY 519 GAGCTGGACAGACTGGAGGAAAGCAGAAAGGAGACGCTACCGCCGCTAGAGCAGCTG 578
Dd 864 AlaGluAsnArgLeuThrGluMetGluThrLeuGlnSerGlnLeuMetAlaGluLysLeu 883
QY 579 CTGCTGGCTGAGAAAGTGTCACAGCGCACCGTGTACGAGCTGGAGAACGAGACACAAG 638
Dd 884 GlnLeuGlnGluGlnLeuGlnAlaGluThrGluLeuCysAlaGluAlaGluGluLeuArg 903
QY 639 CACACTGACTACATGAACAAGAGCGAC-----GACTTCACCAACCTG 680
Dd 904 AlaArgLeuThrAlaLysLysGlnGluLeuGluGluLeuGluCysHisAspLeuGluAlaArg 923
QY 681 CTGGAGCAGGAGCGAGAGAGGTTGAAAAAGCTCCTTGAACAAGAAAAAGCTTACCAAGCC 740
Dd 924 ValGluGluGluGluArgCysGlnHisLeuGlnAlaGluLysLysMetGlnGln 943
QY 741 CGC-----AAAGAAAAAGGAAACGCTAAG-----CGG 767
Dd 944 AsnIleGlnGluGluGluGlnLeuGluGluGluSerAlaArgGlnLysLeuGln 963
QY 768 CTCACAACAACTTCGAGATGAG-----CTTGTGAAGCTCAAGTCCCTTCGCCCTCATGTTG 821
Dd 964 LeuGluLysValThrThrGluAlaLysLeuLysLysLeuGluGluGluGlnIleLeu 983
QY 822 GTGGACGAGAGGCGAGATGCACATCGAGCAACTGGGCTGCAGAGTCAGAAAGTCCAGGAC 881
Dd 984 GluAspGlnAsnCysLysLeuAlaLysGluLysLysLeuGluAspArgIleAlaGlu 1003
QY 882 CTCACTCAGAAGCTGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 941
Dd 1004 PheThrThrAsnLeuThrGluGluGluGluLysSerLysSerLeuAla---LysLeuLys 1022
QY 942 GAAGACCGCCAGAGCTGCTCAAGTTAGAAGTGGACTTCGAACACACAGGCGCTCGAGGTTT 1001
Dd 1023 AsnLysHisGluAlaMetile-----ThrAspLeuGluGluArgLeuArgGlu 1039
QY 1002 TCCAGGAGCACGAAGAGATGAACGCCAAAATTGGCGAATCAAGAATCTCACAAACCGGCAA 1061
Dd 1040 GluLysGlnArgGlnGluLeu-----GluLysThrArgArgLys 1052
QY 1062 CTTGCACTCAAACCTGGTTGGCTTATCGCAAAGGATGAGGAGCTGGAA----- 1109
Dd 1053 LeuGluGlyAspSerThrAspLeuSerAspGlnIleAlaGluLeuGlnAlaGlnIleAla 1072
QY 1110 GAGACCAATAAAAGCCTTCAGAAAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1169
Dd 1073 GluLeuLysMetGlnLeuAlaLysLysGluGluGluGlnAlaLeuAlaLeuAlaArgVal 1092
QY 1170 GCCAAAGGGGAATGTGGAAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGC 1229
Dd 1093 GluGluGluAlaAlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGluSerGln 1112
QY 1230 GTGCTTGAGATGGAG----- 1244
Dd 1113 IleSerGluLeuGlnGluAspLeuGluSerGluArgAlaSerArgAsnLysAlaGluLys 1132
QY 1245 -----GGCAAGGATGAAGAGATCATCGAAGACCGAG----- 1274
Dd 1133 GlnLysArgAspLeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeu 1152
QY 1275 -----GCCAGTGGCGGAGCTGAAG----- 1295
Dd 1153 AspSerThrAlaAlaGlnGlnGluLeuArgSerLysArgGluGlnGluValAsnIleLeu 1172

QY 1296 AAGAAGCTCCAAGAGGAAGAACACACAGCAAGGAACCTTAGACTAGAAAGTGGAGAGCTG 1355
Db 1173 LysLysThrLeuGluGluAlaLysThrHisGluAlaGlnIleGln--GluMetArg 1191
QY 1356 CAGAAGAGGATGTCTGAGCTGGAGAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAA 1415
Db 1192 GlnLysHisSerGlnAlaValGluGluLeuAlaGluGlnLeuGluGlnThrLys----- 1209
QY 1416 TGCACCCAGCTCCATCTGACCTGGAGAAGGAGAAG-----AACCTAACCAAA 1463
Db 1210 -----ArgValLysAlaAsnLeuGluLysAlaLysGlnThrLeuGluAsnGluArgGly 1227
QY 1464 GACCTGCTGAACGAGCTGGAGGTG----- 1487
Db 1228 GluLeuAlaAsnGluValLysValLeuLeuGlnGlyLysGlyAspSerGluHisLysArg 1247
QY 1488 -----GTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAGGCC 1541
Db 1248 LysLysValGluAlaGlnLeuGlnGluLeuGlnValLysPheAsnGluGlyGluArgVal 1267
QY 1542 GAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAGTCTCTCACTGTGATGCTGTGGAT 1601
Db 1268 ArgThrGluLeuAlaAspLysValThrLysLeuGlnVal----- 1280
QY 1602 GAGAGGAAAAATATGATGGAGAAAATAAAGCAAGACAGAGGAAAAGTGGATGGGTGAAT 1661
Db 1281 GluLeuAspAsnValThrGlyLeuLeuSerGlnSerAspSerLysSerLysLeuThr 1300
QY 1662 AAAAAGCTTTAAGGTGGAGCGGAAAAGTCATGGATGTACGGAAAAGCTAATCGAGGAA 1721
Db 1301 LysAspPheSerAlaLeuGluSerGlnLeuGlnAspThrGlnGluLeuGlnGluGlu 1320
QY 1722 AGCAAGAAGCTTTTAAACTCAAATCTGAAATGGAGGAAAAGGAGTACAGTCTGCACAAAG 1781
Db 1321 AsnArgGlnLysLeuSerLeuSerThrLysLeuLysGlnValGlu-----Asp 1336
QY 1782 GAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGGAAGAAGAAAGTCTGTGAACTGAGC 1841
Db 1337 GluLysAsnSerPheArgGluGlnLeuGluGluGluGluAlaLysHisAsnLeuGlu 1356
QY 1842 TGCAGTGTA-----GACTTACTAAAGAACGGCTTGATGGC 1877
Db 1357 LysGlnIleAlaThrLeuHisAlaGlnValAlaAspMetLysLysLysMetGluAspSer 1376
QY 1878 ATA-----GAGGAGGTAGAAAGGGAATAAACCGA-----GGT 1910
Db 1377 ValGlyCysLeuGluThrAlaGluGluValLysArgLysLeuGlnLysAspLeuGluGly 1396
QY 1911 AGTCTGTCAAGGGTCTGAGTTACCTGCCCGGAAGACAATAAGATCAGAGAAGTAACG 1970
Db 1397 LeuSerGlnArgHisGluGluLysValAlaAlaTyrAsp----- 1409
QY 1971 CTTGAAATCGAGAGACTGAAGAAAACGGCTCCAGCAG--TTGGAG-----GTGGTGAG 2021
Db 1410 ---LysLeuGluLysThrLysThrArgLeuGlnGlnGluLeuAspAspLeuLeuValAsp 1428
QY 2022 GGGGAC-----TTGATGAAGACCGAGGACGAAATATGACCAG 2057
Db 1429 LeuAspHisGlnArgGlnSerAlaCysAsnLeuGluLysLysGlnLysLysPheAspGln 1448
QY 2058 TTG-----GAGCAGAAAGTTTCAGAACCGAGCAGGATAGGCAAAAC 2096
Db 1449 LeuLeuAlaGluGluLysThrIleSerAlaLysTyrAlaGluGluArgAspArgAlaGlu 1468
QY 2097 TTCTCTCCAGCAGCTCGAGGAAATCAAAACACCAATGGCCAAGCACAAGCCATAGAG 2156
Db 1469 AlaGluAlaArgGluLysGluThrLysAlaLeuSerLeuAla-----ArgAlaLeuGlu 1486
QY 2157 AAAGGGAGGCGGTAGCCAGGAAGCCGAAGTCCGACAC-----AGGTTTCGGCTG 2207
Db 1487 -----GluAlaMetGluGlnLysAlaGluLeuGluArgLeuAsnLysGlnPheArgThr 1504

QY 2208 GAG-----GAG 2213
Db 1505 GluMetGluAspLeuMetSerSerLysAspAspValGlyLysSerValHisGluLeuGlu 1524
QY 2214 GCTAAAAGTCGTGATTACAGGCCGAGGTGCAGGCTCTCAAGGAGAAGATCCACGAGTGT 2273
Db 1525 LysSerLysArgAlaLeuGluGlnValGluGluMetLysThrGlnLeuGluGluLeu 1544
QY 2274 ATGAACAAG-----GAAGACCAGCTGTCTCAGCTCCAAGTCGACTATTCCGCTC 2321
Db 1545 GluAspGluLeuGlnAlaThrGluAspAlaLysLeuArgLeuGluValAsnLeuGlnAla 1564
QY 2322 -----CTTCAGCAAAAGATTATGGAAGAAGAACTAAGAAC 2357
Db 1565 MetLysAlaGlnPheGluArgAspLeuGlnGlyArgAspGluGlnSerGluGluLysLys 1584
QY 2358 AAGAACATGGGGAGGAGGTCTCAATCTGACCACGAGGAGCTAGAGCTTTCCACGCGCTAC 2417
Db 1585 LysGlnLeuValArgGlnValArgGluMetGluAlaGluLeuGlu---AspGluArgLys 1603
QY 2418 AGCCGAGCTCTCAGCCGAGTGGGAACGGCCGAGGATGGTGGACGTGCCTGTGGCCTCC 2477
Db 1604 GlnArgSerMetAlaValAlaAlaArgLysLysLeuGluMetAspLeuLys----- 1620
QY 2478 ACTGGGTGCAGACCAGCGGTGTGCGGGATGTGCGGAGGAGGAGACCCCGGCTGTG 2537
Db 1621 -----AspLeuGluAlaHisIleAspSerAlaAsnLysAsnArgAspGluAlaIle 1637
QY 2538 TTCATTGCAAAATCTTCAGGAGGAA----- 2564
Db 1638 LysGlnLeuArgLysLeuGlnAlaGlnMetLysAspCysMetArgGluLeuAspAspThr 1657
QY 2565 -----AATCACATCATGAGTAATCTTCACAGGTAGGCCTGAAG---AAACCC 2609
Db 1658 ArgAlaSerArgGluGluIleLeuAlaGlnAlaLysGluAsnGluLysLysLeuLysSer 1677
QY 2610 ATGGAACGGTCTCGTCTCGACAGGTATCCCCAGCAGCGAATGAGCTCACCATGAGG 2669
Db 1678 MetGluAlaGluMetIleGlnLeuGlnGluLeuAlaAlaGluArgAlaLysArg 1697
QY 2670 AAGTCTTGGATTCCTGGATGAGAAAAAGAACCGGTCTTCCACTCCGCAGGAGAAA 2729
Db 1698 GlnAla-----GlnGlnGluArg 1703
QY 2730 GGGCCAGGCCAACCCAGGGTGCAGGGCACCCCGGGAGCTGGTCTCTAGCACCAAGCAG 2789
Db 1704 AspGluLeuAlaAspGluIleAlaAsnSerSerSerGlyLysGlyAlaLeuAlaLeuGluGlu 1723
QY 2790 GGCCAGCCCTACACATCCGTGTG 2813
Db 1724 LysArgArgLeuGluAlaArgIle 1731

RESULT 4

USO1_YEAST
ID USO1_YEAST STANDARD; PRT; 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Intracellular protein transport protein USO1.
GN USO1 OR INT1 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol1, is required for intracellular
protein transport in Saccharomyces cerevisiae.";

DR InterPro; IPR004009; Myosin N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Multigene family.
FT DOMAIN 1 778 MYOSIN HEAD-LIKE.
FT DOMAIN 779 808 IQ.
FT DOMAIN 841 1927 COILED COIL (POTENTIAL).
FT NP BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 654 676 ACTIN-BINDING.
SQ SEQUENCE 1961 AA; 226336 MW; 9B9876D9681FB19E CRC64;

Alignment Scores:
Pred. No.: 7.92e-12 Length: 1961
Score: 458.50 Matches: 195
Percent Similarity: 43.13% Conservative: 163
Best Local Similarity: 23.49% Mismatches: 283
Query Match: 5.93% Indels: 189
DB: 1 Gaps: 33

US-10-788-793-1 (1-4364) x MYH9_RAT (1-1961)
QY 447 GTCCTGCACCGAGATGCCATCCTTGCTCAAGAGAAGTCCATAGGAGAGACGTCCTATGAG 506
Db 841 ileArgHisGluAspGluLeuLeuAlaLysGluAlaGluLeuThrLys---ValArgGlu 859
QY 507 AAACCTATCTCAGAGCTGGACAGACTGGAGGAAAGCAGAAGGAGACGTACCGCCGCAATG 566
Db 860 LysHisLeuAlaAlaGluAsnArgLeuThrGluMetGluThrMetGlnSerGlnLeuMet 879
QY 567 CTAGAGCAGCTGCTGCTGGCTGAGAAAGTGTCACAGGCGCACCGTGTACGAGCTGGAGAAC 626
Db 880 AlaGluLysLeuGlnLeuGlnGluGlnLeuGlnAlaLysThrGluLeuCysAlaGluAla 899
QY 627 GAGAAGCACAAAGCACACTGACTACATGAAACAAGAGCGAC-----GAC 668
Db 900 GluGluLeuArgAlaArgLeuThrAlaLysLysGlnGluLeuGluGluLeuCysHisAsp 919
QY 669 TTCACCAACCTGCTGGACGAGCGGAGAGAGGTTGAAAAAGCTCCTTGAACAAGAAAAA 728
Db 920 LeuGluAlaArgValGluGluGluGluArgCysGlnTyrLeuGlnAlaGluLysLys 939
QY 729 GCTTACCAGCCCGC-----AAAGAAAGGAAACCGCTAAG 764
Db 940 LysMetGlnGlnAsnIleGlnGluLeuGluGlnLeuGluGluGluGluSerAlaArg 959
QY 765 -----CGGCTCAACAAACTTCGAGATGAG-----CTTGTGAAGCTCAAGTCCTTC 809
Db 960 GlnLysLeuGlnLeuGluLysValThrThrGluAlaLysLeuLysLysLeuGluGluAsp 979
QY 810 GCCCTCATGTTGGTGACGAGAGGCGAGATGACATCGAGCAACTGGGCTGCGAGAGTCAG 869
Db 980 GlnIleIleMetGluAspGlnAsnCysLysLysLeuAlaLysGluLysLysLeuGluAsp 999
QY 870 AAAGTCAGGACCTCACTCAGAAGCTGAGGGAGGAGGAGAAAGAACTCAAAGCGGTCACT 929
Db 1000 ArgValAlaGluPheThrThrAspLeuMetGluGluGluLysSerLysSerLeuAla 1019
QY 930 TACAAATCAAGGAAGACCGCCAGAAAGCTGCTCAAGTTAGAAAGTGGAAGTTCGAACACAAAG 989
Db 1020 ---LysLeuLysAsnLysHisGluAlaMetIle-----ThrAspLeuGluGluArg 1035
QY 990 GCCTCGAGGTTTCCGAGGACGACGAAGAGATGAACGCCAAATTTGGCGAATCAAGAACTCT 1049
Db 1036 LeuArgArgGluGluLysGlnArgGlnGluLeu-----GluLys 1048

QY 1050 CACAACCGGCAACTTCGACTCAAACTGGTGGCTTATCGCAAAAGGATTCAGGAGCTGGAA 1109
Db 1049 ThrArgArgLysLeuGluGlyAspSerThrAspLeuSerAspGlnIleAlaGluLeuGln 1068
QY 1110 -----GAGACCAATAAAAGCCTTCAGAGCGCAGAGGAAAGCTCCAGGAGCTG 1157
Db 1069 AlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGluLeuGlnAlaAla 1088
QY 1158 AGAGAGAAATTCCTCAAGGGGAATCTGGAACTCCAGTCTCATGGCGGAAGTGGAGAGT 1217
Db 1089 LeuAlaArgValGluGluAlaAlaGlnLysAsnMetAlaLeuLysLysIleArgGlu 1108
QY 1218 CTGCGCAAGCGCTGCTTGAGATGGAG----- 1244
Db 1109 LeuGluThrGlnIleSerGluLeuGlnGluAspLeuGluSerGluArgAlaCysArgAsn 1128
QY 1245 -----GGCAAGGATGAAGAGATCACGAAGACCGAG--- 1274
Db 1129 LysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuAlaLeuLysThrGluLeu 1148
QY 1275 -----GCCCAGTCCGGGAGCTGAAGAAGAAAGCTCCAAGAGGAA 1313
Db 1149 GluAspThrLeuAspSerThrAlaAlaGlnGlnGluLeuArgSerLysArgGluGlnGlu 1168
QY 1314 -----GAACACCACACGCAAGGAACCTTAGACTAGAAAGTGGAG 1349
Db 1169 ValSerIleLeuLysLysThrLeuGluAspGluAlaLysThrHisGluAlaGlnIleGln 1188
QY 1350 AAGCTG--CAGAAGAGGATGTCTGAGCTGGAGAAGCTGGAGGAAGCGTTCAGCCGGAGT 1406
Db 1189 GluMetArgGlnLysHisSerGlnAlaValGluGluLeuAlaGluGlnLeuGluGlnThr 1208
QY 1407 AAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAGGAGAAG-----AAC 1454
Db 1209 Lys-----ArgValLysAlaThrLeuGluLysAlaLysGlnThrLeuGluAsn 1224
QY 1455 CTAACCAAGACCTGCTGAACGAGCTGGAGGTG----- 1487
Db 1225 GluArgGlyGluLeuAlaAsnGluValLysAlaLeuLeuGlnGlyLysGlyAspSerGlu 1244
QY 1488 -----GTCAAGAGTCGAGTTAAAGAACTCGAATGTCTCCGAGAGTAGACTG 1532
Db 1245 HisLysArgLysLysValGluAlaGlnLeuGlnGluLeuValLysPheSerGluGly 1264
QY 1533 GAGAAGCCCGAGTTAAGCCTCAAAGATGACCTTACAAGCTGAAGTCCCTCCTGCTGATG 1592
Db 1265 GluArgValArgThrGluLeuAlaAspLysValSerLysLeuGlnVal----- 1280
QY 1593 CTGCTGATGAGAGGAAAAATATGATGAGAGAAAAATAAAACAAGAAAGAGAGAAAGTGGAT 1652
Db 1281 -----GluLeuAspSerValThrGlyLeuLeuAsnGlnSerAspSerLysSerSer 1297
QY 1653 GGGTTGAATAAAACTTTAAGGTGGAGCGGGAAGATCATGGATGTGACGGAAAAAGCTA 1712
Db 1298 LysLeuThrLysAspPheSerAlaLeuGluSerGlnLeuGlnAspThrGlnGluLeuLeu 1317
QY 1713 ATCGAGGAAAGCAAGAGCTTTTAAACTCAAATCTGAAATGGAGGAAAAAGGAGTACAGT 1772
Db 1318 GlnGluGluAsnArgGlnLysLeuSerLeuSerThrLysLeuLysGlnMetGlu----- 1335
QY 1773 CTGACAAAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGCAAGAGAA----- 1823
Db 1336 -----AspGluLysAsnSerPheArgGluGlnLeuGluGluGluGluAlaLys 1353
QY 1824 AGGTCCTGTGAACCTGAGCTGCAGT-----GTAGACTTACTAAAGAAG 1865
Db 1354 ArgAsnLeuGluLysGlnIleAlaThrLeuHisAlaGlnValThrAspMetLysLysLys 1373
QY 1866 CGGCTTCATGGCATA-----GAGGAGGTAGAAAGGGGAAATAAACCGA 1907
Db 1374 MetGluAspGlyValGlyCysLeuGluThrAlaGluGluAlaLysArgArgLeuGlnLys 1393

QY 1908 -----GGTAGGTCGTCAAGGGGTCTGAGTTACCTGCCCGGAAGACAATAAGATC 1958
Db 1394 AspLeuGluGlyLeuSerGlnArgLeuGluGlyLysValAlaAlaTyrAsp----- 1410
QY 1959 AGAGAACTAACGGTTGAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGTGGTG 2018
Db 1411 -----LysLeuGluLysThrLysThrArgLeuGlnGln---GluLeuAsp 1424
QY 2019 GAG-----GGGACTTGATGAGACCGAG 2042
Db 1425 AspLeuLeuValAspLeuAspHisGlnArgGlnSerValSerAsnLeuGluLysGln 1444
QY 2043 GACGAATATGACCAGTTG-----GAGCAGAAGTTCAGAACCGAG 2081
Db 1445 LysLysPheAspGlnLeuLeuAlaGluGluLysThrIleSerAlaLysTyrAlaGluGlu 1464
QY 2082 CAGGATAAGGCAAACTTCTCTCCAGCAGCTCGAGGAATCAACACCAATGCCCAAG 2141
Db 1465 ArgAspArgAlaGluAlaGluAlaArgGluLysGluThrLysAlaLeuSerLeuAla--- 1483
QY 2142 CACAAAGCCATAGAGAAAGGGAGGCGGTGAGCCAGGAAAGCCGAACTGCGACAC----- 2195
Db 1484 ---ArgAlaLeuGlu-----GluAlaMetGluGlnLysAlaGluLeuGluArgLeuAsn 1500
QY 2196 ---AGGTTTCGGCTGGAG----- 2210
Db 1501 LysGlnPheArgThrGluMetGluAspLeuMetSerLysAspValGlyLysSer 1520
QY 2211 -----GAGGCTAAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAGGAG 2258
Db 1521 ValHisGluLeuGluLysSerAsnArgAlaLeuGluGlnValGluGluMetLysThr 1540
QY 2259 AAGATCCACGAGCTGATGAACAAG-----GAAGACCAGCTGTCTCAGCTCCAA 2306
Db 1541 GlnLeuGluGluLeuGluAspGluLeuGlnAlaThrGluAspAlaLysLeuArgLeuGlu 1560
QY 2307 GTCGACTATTCCGGTC-----CTTCAGCAAAAGATTATGGAA 2342
Db 1561 ValAsnLeuGlnAlaMetLysAlaGlnPheGluArgAspLeuGlnGlyArgAspGluGln 1580
QY 2343 GAAGAAACTAAGAACAAAGAACATGGGAGGAGGTCTCAATCTGACCAAGGAGCTAGAG 2402
Db 1581 SerGluGluLysLysGlnLeuValArgGlnValArgGluMetGluAlaGluLeuGlu 1600
QY 2403 CTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAGGATGGTGAC 2462
Db 1601 ---AspGluArgLysGlnArgSerIleAlaMetAlaAlaArgLysLysLeuGluMetAsp 1619
QY 2463 GTGCCTGTGGCCTCCACTGGGGTGACACC 2492
Db 1620 LeuLysAspLeuGluAlaHisIleAspThr 1629

RESULT 6

MYS_AEQIR
ID_MYS_AEQIR STANDARD; PRT; 1938 AA.
AC P24733;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, striated muscle.
OS Aequipecten irradians (Bay scallop).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pectinoidea;
OC Pectinoidea; Pectinidae; Argopecten.
OX NCBI_TaxID=31199;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Adductor muscle;
RX MEDLINE=92011595; PubMed=1917970;
RA Nyitray L., Goodwin E.B., Szent-Gyorgyi A.G.;
RT "Complete primary structure of a scallop striated muscle myosin heavy
chain. Sequence comparison with other heavy chains reveals regions
that might be critical for regulation."
RL J. Biol. Chem. 266:18469-18476(1991).

RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Adductor muscle;
RX MEDLINE=91088319; PubMed=2263488;
RA Nyitray L., Goodwin E.B., Szent-Gyorgyi A.G.;
RT "Nucleotide sequence of full length cDNA for a scallop striated
muscle myosin heavy chain."
RL Nucleic Acids Res. 18:7158-7158(1990).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 777-836.
RX MEDLINE=94173332; PubMed=8127365;
RA Xie X., Harrison D.H., Schlichting I., Sweet R.M., Kalabokis V.N.,
Szent-Gyorgyi A.G., Cohen C.;
RT "Structure of the regulatory domain of scallop myosin at 2.8-A
resolution."
RL Nature 368:306-312(1994).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 777-836.
RX MEDLINE=96419133; PubMed=8805510;
RA Houdusse A., Cohen C.;
RT "Structure of the regulatory domain of scallop myosin at 2-A
resolution: implications for regulation."
RL Structure 4:21-32(1996).
CC -|- FUNCTION: Muscle contraction.
CC -|- FUNCTION: Myosin is a protein that binds to F-actin and has ATPase
activity that is activated by F-actin.
CC -|- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
heavy chain subunits (MHC), 2 alkali light chain subunits (MLC-2)
and 2 regulatory light chain subunits (MLC-2).
CC -|- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -|- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -|- SIMILARITY: Contains 1 IQ domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X55714; CAA39247.1; -.
CC PIR; A40997; A40997.
CC PDB; 1SCM; 30-APR-94.
CC PDB; 1WDC; 11-JUL-96.
CC PDB; 1B7T; 12-MAY-99.
CC PDB; 1KK7; 09-OCT-02.
CC PDB; 1KQM; 20-NOV-02.
CC PDB; 1L2O; 30-OCT-02.
CC InterPro; IPR000048; IQ region.
CC InterPro; IPR001609; myosin head.
CC InterPro; IPR004009; Myosin_N.
CC InterPro; IPR002928; Myosin_tail.
CC InterPro; IPR002017; Spectrin.
CC Pfam; PF00612; IQ; 2.
CC Pfam; PF00063; myosin head; 1.
CC Pfam; PF02736; Myosin_N; 1.
CC Pfam; PF01576; Myosin_tail; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00015; IQ; 1.
CC SMART; SM00242; MYSC; 1.
CC PROSITE; PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Calmodulin-binding; 3D-structure.
FT DOMAIN 1 777 MYOSIN HEAD-LIKE.
FT DOMAIN 778 805 IQ.
FT DOMAIN 836 1938 RODLIKE TAIL (S2 AND LMM DOMAINS).
FT DOMAIN 836 1938 COILED COIL (POTENTIAL).
FT NP_BIND 176 183 ATP (BY SIMILARITY).
FT HELIX 776 822
FT TURN 823 823


```
CC CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC CC -!- SIMILARITY: Contains 1 IQ domain.
CC CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC CC WWW="http://www.infobiogen.fr/services/chroncancer/Genes/MYH11ID43.html".
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; AF001548; AAC31665.1; -.
CC CC EMBL; U91323; AAC35212.1; -.
CC CC EMBL; AB020673; BAA74889.2; ALT INIT.
CC CC EMBL; D10667; -; NOT ANNOTATED_CDS.
CC CC EMBL; X69292; CAA49154.1; -.
CC CC HSSP; P10587; 1BR2.
CC CC Genew; HGNC:7569; MYH11.
CC CC MIM; 160745; -.
CC CC GO; GO:0005859; C:muscle myosin; TAS.
CC CC InterPro; IPR000048; IQ_region.
CC CC InterPro; IPR001609; myosin_head.
CC CC InterPro; IPR004009; Myosin_N.
CC CC InterPro; IPR002928; Myosin_tail.
CC CC InterPro; IPR002017; Spectrin.
CC CC Pfam; PF00612; IQ; 1.
CC CC Pfam; PF00063; myosin_head; 1.
CC CC Pfam; PF02736; Myosin_N; 1.
CC CC Pfam; PF01576; Myosin_tail; 1.
CC CC PRINTS; PR00193; MYOSINHEAVY.
CC CC ProDom; PD000355; myosin_head; 1.
CC CC SMART; SM00015; IQ; 1.
CC CC SMART; SM00242; MYSC; 1.
CC CC PROSITE; PS50096; IQ; 1.
CC CC KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
CC KW Calmodulin-binding; ATP-binding; Methylation; Multigene family;
CC KW Proto-oncogene; Chromosomal translocation.
CC FT DOMAIN 1 785 MYOSIN HEAD-LIKE.
CC FT DOMAIN 786 815 IQ.
CC FT DOMAIN 844 1934 COILED COIL (POTENTIAL).
CC FT DOMAIN 1935 1972 CARBOXYL-TERMINAL.
CC FT NP_BIND 178 185 ATP (POTENTIAL).
CC FT DOMAIN 661 683 ACTIN-BINDING (BY SIMILARITY).
CC FT DOMAIN 762 776 ACTIN-BINDING (BY SIMILARITY).
CC FT MOD_RES 129 129 METHYLATION (TRI-) (POTENTIAL).
CC FT CONFLICT 887 889 EEK -> NSE (IN REF. 4).
CC FT CONFLICT 1558 1558 T -> S (IN REF. 4).
CC FT CONFLICT 1610 1611 KQ -> NE (IN REF. 4).
CC FT CONFLICT 1786 1786 A -> S (IN REF. 5).
CC FT CONFLICT 1958 1958 T -> L (IN REF. 4).
CC SQ SEQUENCE 1972 AA; 227338 MW; 67665BBB2AECE1277 CRC64;

Alignment Scores:
Pred. No.: 1.44e-11 Length: 1972
Score: 452.00 Matches: 195
Percent Similarity: 43.73% Conservative: 168
Best Local Similarity: 23.49% Mismatches: 301
Query Match: 5.85% Indels: 166
DB: 1 Gaps: 29

US-10-788-793-1 (1-4364) x MYHB_HUMAN (1-1972)
QY 273 AAGACTAAGAAGTCTGTGGAGTTATCCAGGAGGACCTCATCCAGCTCCTGAGTATCATG 332
Dbb ||| ||| ::::::::::::::::::::
Dbb 840 LysValLysProLeuLeuGlnValThrArgGln----- 850
QY 333 GAAGGGGAGTTGCAGGTCGAGAAAGATGTTCATCCACATGCTGAGGACAGAGAAACCAAG 392
Dbb ||| ||| ::::::::::::::::::::
Dbb 851 GluGluGluMetGlnAlaLysGluAspGluLeu-----GlnLysThrLys 865
QY 393 CCCGAGGTTCTGGAGGCACACTATGGATCTGCAGAACCTTGAGAAAGTGTCCGGTCCCTG 452
```

```
Dbb 866 GluArgGlnGlnLysAlaGluAsnGluLeuLysGluLeuGln----- 880
QY 453 CACCGAGATGCCATCCTTGCTCAAGAGAAGTCCATGAGAGAAGACGTCTATGAGAAAACCT 512
Dbb :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 513 ATCTCAGAGCTGGACAGACTGGAGGAAAAGCAGAGAGACGTACCGCCGCTAGAG 572
Dbb :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 573 CAGCTGCTGCTGGCTGAGAAGTGTACACAGGCGCACCCGTGTACGAGCTGGAGAACGAGAAG 632
Dbb ||| ||| :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 633 CACAAGCACACTGACTACATGAACAAGAGCGACGACTTCACCAACCTGCTGGAGCAGGAG 692
Dbb ||| :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 693 CGAGAGAGG-----TTGAAAAAGCTCCTTGAAACAAGAAAAAGCT 731
Dbb ||| :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 732 TACCAAGCCCGCAAGAAAAAGGAAACGCT-----AAGCGGCTCAACAAACTTCGA 782
Dbb :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 783 GATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCATGTTGGTGGACGAGGCGCAGATGCAC 842
Dbb ||| :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 843 ATCGAGCAACTGGCGCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAAGCTGAGGGAG 902
Dbb :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 903 GAGGAAGAAAAACTCAAAGCGGTCACT----- 929
Dbb ||| ||| ||| :::: :::: :::: :::: :::: :::: :::: :::: ::::
Dbb 1018 GluGluLysAlaLysAsnLeuThrLysLeuLysAsnLysHisGluSerMetIleSer 1037
QY 930 -----TACAAATCCAAGGAAGACCGCCAGAGCTGCTCAAGTTAGAA 971
Dbb ||| :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Dbb 1038 GluLeuGluValArgLeuLysLysGluLysSerArgGlnGluLysLeuLys 1057
QY 972 GTGGACTTCGAACACAAAGGCTCGAGGTTTCCCAGGACGACGAGAGATGAACGCCAAA 1031
Dbb ||| ||| ||| :::: :::: :::: :::: :::: :::: :::: ::::
Dbb 1058 ArgLysLeuGluGlyAspAlaSerAppPheHisGluGlnIleAlaAspLeuGlnAlaGln 1077
QY 1032 TTGGCGAATCAAGAA-----TCTCACAAACCGGCAACTTCGACTCAAACCTGGTT 1079
Dbb :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Dbb 1078 IleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluLeuGlnAlaAlaLeuAla 1097
QY 1080 GGCTTATCGCAAGGATTGAGGAGCTGGAAGAGACCAATAAAAAGCCTTCAGAAGGCAGAG 1139
Dbb ||| ||| :::: :::: :::: :::: :::: :::: :::: :::: ::::
Dbb 1098 ArgLeuAspAspGluIleAlaGlnLysAsnAlaLeuLysLysIleArgGluLeuGlu 1117
QY 1140 GAAGAGCTCCAGGAGCTGAGAGAGAAAATT-----GCCAAAGGG 1178
Dbb :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Dbb 1118 GlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArgAlaAlaArgAsnLysAla 1137
QY 1179 GAATGTGGAACCTCCAGTCTCATGGCGGAAAGTGGAGAGTCTGCGCAAGCGCGTG----- 1232
Dbb ||| ||| ||| :::: :::: :::: :::: :::: :::: :::: ::::
Dbb 1138 GluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAsp 1157
QY 1233 ---CTTGAGATGGAGGCAAGGATGAAGAGATC---ACGAAGACCGAGGCCCGCCAGTCCCG 1286
Dbb ||| :::: :::: :::: :::: :::: :::: :::: :::: ::::
Dbb 1158 ThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLysArgGluGlnGluValThr 1177
QY 1287 GAGCTGAAGAAGAGCTCCAAGAGGAA---GAACACCACAGCAAGGAACCTTAGACTAGAA 1343
Dbb ||| ||| ||| ||| ||| :::: :::: :::: :::: :::: ::::
Dbb 1178 ValLeuLysLysAlaLeuAspGluGluThrArgSerHisGluAlaGlnValGlnGluMet 1197
QY 1344 GTGGAGAAGCTGCAGAAGAGGATGTCTGAGCTGGAGAGCTGGAGGAAGCGTTTCAGCCGG 1403
Dbb :::: ||| :::: :::: :::: :::: :::: :::: ||| ||| |||
```

Db 1198 ArgGlnLysHisAlaGlnAlaValGluGluLeuThrGluGlnLeuGluGlnPheLysArg 1217
QY 1404 AGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAGGAGAAGAACCTAACCAAA 1463
Db 1218 AlaLysAlaAsnLeuAspLysAsnLysGlnThrLeuGluLysGlu-----AsnAla 1234
QY 1464 GACCTGCTGAACGAGCTGGAGGTG----- 1487
Db 1235 AspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGluHisLysLys 1254
QY 1488 -----GTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAGGCC 1541
Db 1255 LysLysLeuGluAlaGlnValGlnGluLeuGlnSerLysCysSerAspGlyGluArgAla 1274
QY 1542 GAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAGTCTTCACTGTGATGCTGGTGGAT 1601
Db 1275 ArgAlaGluLeuAsnAspLysValHisLysLeuGln-----Asn 1287
QY 1602 GAGAGGAAAAATATGATGGAGAAAATAAAGCAAGACAGAGGAAAAGTGGATGGGTGTAAT 1661
Db 1288 GluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLysAlaIleLysLeuAla 1307
QY 1662 AAAAACTTTAAGGTGGAGCAGGGAAAAAGTCATGGATGTACGGAAAAAGCTAATCGAGGAA 1721
Db 1308 LysAspValAlaSerLeuSerGlnLeuGlnAspThrGlnGluLeuGlnGluGlu 1327
QY 1722 AGCAAGAAGCTTTTAAACTCAAATCT-----GAAATCGAGGAAAAGGAGTACAGT 1772
Db 1328 ThrArgGlnLysLeuAsnValSerThrLysLeuArgGlnLeuGluGluArgAsnSer 1347
QY 1773 CTGACAAAGGAGGAGGTGAGCTGATGGGTAAACTGAGGAGCGGAAGAAGAGTCTCTGT 1832
Db 1348 LeuGlnAspGlnLeuAspGluGluMetGluAlaLysGlnAsnLeuGluArgHisIleSer 1367
QY 1833 GAACTGAGCTGCAGTGTAGACTTACTAAAGAAGCGGCTT----- 1871
Db 1368 ThrLeuAsnIleGlnLeuSerAspSerLysLysLysLeuGlnAspPheAlaSerThrVal 1387
QY 1872 GATGGCATAGAGGAG----- 1886
Db 1388 GluAlaLeuGluGluGlyLysLysArgPheGlnLysGluIleGluAsnLeuThrGlnGln 1407
QY 1887 -----GTAGAAAGGGAAATAAACCCGAGGTAGGTCTG 1916
Db 1408 TyrGluGluLysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGln 1427
QY 1917 TGCAAGGGGTCTGAGTTACCTGCCCCGGAAGACATAAGATCAGAGAACTAACGCTTGAA 1976
Db 1428 ---GluLeuAspAspLeuValValAspLeuAspAsnGln---ArgGlnLeuValSerAsn 1445
QY 1977 ATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGTGGTGGAGGGGACTTG---ATG 2033
Db 1446 LeuGluLysLysGlnArgLysPheAspGlnLeuLeuAlaGluGluLysAsnIleSerSer 1465
QY 2034 AAGACCGAGGACGAATATGACAGTTGGAGCAGAGTTTCAGAACCGAGCAGGATAAGGCA 2093
Db 1466 LysTyrAlaAspGluArgAspArgAlaGluAlaGluAlaArgGluLysGluThrLysAla 1485
QY 2094 AACTTCCTCTCCAGCAGCTCGAG-----GAAATCAAAACACCAAAATGGCCCAAGCAC 2144
Db 1486 LeuSerLeuAlaArgAlaLeuGluGluAlaLeuGluAlaLysGluGluLeuGluArgThr 1505
QY 2145 AAAGCCATAGAGAAAGGGGAG-----GCCGTGACCCAGGAAGCCGAACCTGCGACAC 2195
Db 1506 AsnLysMetLeuLysAlaGluMetGluAspLeuValSerSerLysAspAspValGlyLys 1525
QY 2196 AGTTTCGGCTGGAGGAGGCTAAAAGTCGTGATTTCAGGCCGAGGTGCAGGCTCTCAAG 2255
Db 1526 AsnValHisGluLeuGluLysSerLysArgAlaLeuGluThrGlnMetGluGluMetLys 1545
QY 2256 GAGAAGATCCACGAGCTGATGAACAAG-----GAAGACCAGCTGTCTCAGCTC 2303
Db 1546 ThrGlnLeuGluGluLeuGluAspGluLeuGlnAlaThrGluAspAlaLysLeuArgLeu 1565

QY 2304 CAAGTCGACTATTTCGGTC-----CTTCAGCAAAAGATTATTATG 2339
Db 1566 GluValAsnMetGlnAlaLeuLysGlyGlnPheGluArgAspLeuGlnAlaArgAspGlu 1585
QY 2340 GAAGAAGAACTAAGAACAAACATGGGAGGAGGAGTCTCAATCTGACCAAGGAGCTA 2399
Db 1586 GlnAsnGluGluLysArgGlnLeuGlnArgGlnLeuHisGluTyrGluThrGluLeu 1605
QY 2400 GAGCTTCCAAAGCGCTACAGCCGAGCTCTC 2429
Db 1606 Glu---AspGluArgLysGlnArgAlaLeu 1614
RESULT 8
MYHB_MOUSE
ID MYHB_MOUSE STANDARD; PRT; 1972 AA.
AC O08638; O08639; Q62462; Q64195;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, smooth muscle isoform (SMMHC).
GN MYH11.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Uterus;
RX MEDLINE=97242182; PubMed=9125171;
RA Hasegawa K., Arakawa E., Oda S., Matsuda Y.;
RT "Molecular cloning and expression of murine smooth muscle myosin heavy
RT chains.";
RL Biochem. Biophys. Res. Commun. 232:313-316(1997).
RN [2]
RP SEQUENCE OF 1-368 FROM N.A.
RC TISSUE=Uterus;
RX MEDLINE=95008063; PubMed=79233625;
RA Miano J.M., Cserjesi P., Ligon K.L., Periasamy M., Olson E.N.;
RT "Smooth muscle myosin heavy chain exclusively marks the smooth muscle
RT lineage during mouse embryogenesis.";
RL Circ. Res. 75:803-812(1994).
RN [3]
RP SEQUENCE OF 1-126 FROM N.A.
RC TISSUE=Uterus;
RX MEDLINE=96172919; PubMed=8593698;
RA Suzuki T., Kim H.S., Kurabayashi M., Hamada H., Fujii H., Aikawa M.,
RA Watanabe M., Watanabe N., Sakomura Y., Yazaki Y., Nagai R.;
RT "Preferential differentiation of P19 mouse embryonal carcinoma cells
RT into smooth muscle cells. Use of retinoic acid and antisense against
RT the central nervous system-specific POU transcription factor Brn-2.";
RL Circ. Res. 78:395-404(1996).
CC -!- FUNCTION: Muscle contraction.
CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
CC and 2 regulatory light chain subunits (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=O08638-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O08638-2; Sequence=VSP_003346;
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
CC meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be
CC split further into 2 globular subfragments (S1) and 1 rod-shaped
CC subfragment (S2).
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.

QY	1365	ATG	TCT	GAG	CTG	GAG	AAAG	CTG	GAG	GAAG	CGG	TTT	CAG	CCG	GAG	TAA	AGT	CGG	AAT	GC	ACC	CAG	1422
		:::		:::		:::		:::		:::		:::		:::		:::		:::		:::		:::	
Db	1205	Val	Glu	Glu	Leu	Thr	Glu	Gln	Leu	Glu	Gln	Phe	Lys	Arg	Ala	Lys	Ala	Asn	Leu	Asp	Lys	1224	
		:::		:::		:::		:::		:::		:::		:::		:::		:::		:::		:::	
QY	1425	CTC	CAT	CTG	AAC	CTG	GAG	GAAG	GAG	AAAC	CTT	AAC	CAAG	ACCT	GCT	GCT	GAAC	GAG	CTG	GAG	1484		
Db	1225	Ser	Lys	Gln	Thr	Leu	Glu	Lys	Glu	-----	Asn	Ala	Asp	Leu	Ala	Gly	Glu	Leu	Arg	1241			
QY	1485	GTG	-----	-----	-----	GT	CA	GAG	TC	GAG	TAA	AA	GA	ACT	CG	AA	-----	-----	-----	-----	1514		
Db	1242	Val	Leu	Gly	Gln	Ala	Lys	Gln	Val	Glu	His	Lys	Lys	Lys	Leu	Glu	Val	Gln	Leu	1261			
QY	1515	-----	-----	-----	-----	TG	CT	CC	GAG	AG	TAG	ACT	TG	GAG	AA	GGC	CG	GAG	TTA	AG	CCT	CAAA	1556
Db	1262	Gln	Asp	Leu	Gln	Ser	Lys	Cys	Ser	Asp	Gly	Glu	Arg	Ala	Arg	Ala	Glu	Leu	Ser	-----	-----	1279	
QY	1557	GAT	GAC	CTT	TACA	AGCT	GAAG	TCCT	TCA	CTG	TGAT	GCT	GGT	GGAT	GAT	GAG	AGG	AAAA	TAAT	G	1616		
Db	1280	Asp	Lys	Val	His	Lys	Leu	Gln	-----	-----	-----	-----	-----	-----	Asn	Glu	Val	Glu	Ser	Val	1292		
QY	1617	ATG	GAG	AAAA	TAA	AGCA	AGAG	GAG	AGG	AAAG	TGG	ATG	GGT	TGA	ATA	AAAA	CTT	TA	AGG	TG	1676		
Db	1293	Thr	Gly	Met	Leu	Asn	Glu	Ala	Glu	Gly	Lys	Ala	Ile	Lys	Leu	Ala	Lys	Asp	Val	Ala	Ser	1312	
QY	1677	GAG	CAG	GGA	AAAG	TCA	TGG	ATG	TGAC	GGA	AAAG	CTA	AT	CG	AGG	AAAG	CAAG	CAAG	CTT	TA	1736		
Db	1313	Leu	Gly	Ser	Gln	Leu	Gln	Asp	Thr														

```

QY 2271 -----CTGATGAACAAGGAGACCAAGCTGTCTCAGCTCCAAGTCGACTATTTCGGTC 2321
Db 1552 GluAspValGlnAlaThrGluAspAlaLysLeuArgLeuGluValAsnMetGlnAla 1571
QY 2322 -----CTTCAGCAAAAGATTATGGAAGAGAAAGAACTAAGAAC 2357
Db 1572 LeuLysGlyGlnPheGluArgAspLeuGlnAlaArgAspGluGlnAsnGluGluLysArg 1591
QY 2358 AAGAACATGGGGAGGAGGCTCCTCAATCTGACCAAGGAGCTAGAGCTTCCAAAGCGCTAC 2417
Db 1592 ArgGlnLeuGlnArgGlnLeuHisGluTyrGluThrGluLeuGlu---AspGluArgLys 1610
QY 2418 AGCCGAGCTCTC 2429
Db 1611 GlnArgAlaLeu 1614

RESULT 9
MYH9_CHICK
ID MYH9 CHICK STANDARD; PRT; 1959 AA.
AC P14105;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, nonmuscle (Cellular myosin heavy chain) (NMHC).
GN MYH9.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestinal epithelium;
RX MEDLINE=90046668; PubMed=2813355;
RA Shohet R.V., Conti M.A., Kawamoto S., Preston Y.A., Brill D.A.,
RA Adelstein R.S.;
RT "Cloning of the cDNA encoding the myosin heavy chain of a vertebrate
cellular myosin.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:7726-7730(1989).
CC -!- FUNCTION: Cellular myosin appears to play a role in cytokinesis,
CC cell shape, and specialized functions such as secretion and
CC capping.
CC -!- SUBUNIT: Myosin is a hexameric protein that consists of 2 heavy
CC chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2
CC regulatory light chain subunits (MLC-2).
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M26510; AAA48974.1; -.
DR PIR; A33977; A33977.
DR HSSP; P10587; 1BR2.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PD00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.

```

DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Multigene family.
FT DOMAIN 1 778 MYOSIN HEAD-LIKE.
FT DOMAIN 779 808 IQ.
FT DOMAIN 837 1925 COILED COIL (POTENTIAL).
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 654 676 ACTIN-BINDING.
SQ SEQUENCE 1959 AA; 226502 MW; A75C86086FD3A1A1 CRC64;

Alignment Scores:
Pred. No.: 2.19e-11 Length: 1959
Score: 447.50 Matches: 224
Percent Similarity: 41.31% Conservative: 192
Best Local Similarity: 22.24% Mismatches: 344
Query Match: 5.79% Indels: 247
DB: 1 Gaps: 41

US-10-788-793-1 (1-4364) x MYH9_CHICK (1-1959)

QY 192 AAGGCCAATCGGAAGGAGGAGGATGTCATGGCTTCCGGAACCTATCAAAAAGGCACCTCAA 251
Db :::::::::::::::::::::
QY 839 GlnValSerArgGlnGluGluMetMetAla 849
Db :::::::::::::::::::::
QY 252 CCATCTGGAGAAAGTGAGAAAAGACTAAGAAGTCTGTGGAGTTATCCCAAGGAGCACCTC 311
Db :::::::::::::::::::::
QY 850 852
QY 312 ATCCAGCTCCTGAGTATCATGGAAGGGAGTTGCAGGCTCGAGAAGATGTCATCCACATG 371
Db :::::::::::::::::::::
QY 853 ---GluLeuIleLysValLysGluLysGlnLeuAlaAlaGluAsnArgLeuSerGluMet 871
QY 372 CTGAGGACAGAGAAAACCAAGCCCGAGGTTCTGGAGGACACTATGGATCTGCAGAACCT 431
Db :::::::::::::::::::::
QY 872 ---GluThrPheGlnAlaGlnLeu---MetAlaGluLys 882
QY 432 GAGAAAGTGCTTCGGGTCTCGCACCAGATGCCATCCTTGCTCAAGAGAAGTCCATAGGA 491
Db :::::::::::::::::::::
QY 883 MetGlnLeuGlnGluGlnLeuGlnAlaGluAlaGluCysAlaGlu-----Ala 899
QY 492 GAAGACGTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAAAGCAGAAGGAG 551
Db :::::::::::::::::::::
QY 900 GluGluIleArgAlaArgLeuThrAlaLysLysGlnGluLeuGluGluIleCysHisAsp 919
QY 552 ACGTACCGCGCATGCTAGACAGCTGCTGCTGGCTGAGAAGTGTACAGCGCACCGTG 611
Db :::::::::::::::::::::
QY 920 LeuGluAlaArgValGluGlu-----GluGluArgCysGln----- 932
QY 612 TACGAGCTGGAGAACGAGACACAAAGCACACTGACTACATGAACAAGAGCGCAGCTTC 671
Db :::::::::::::::::::::
QY 933 ---HisLeuGlnAlaGluLysLys 940
QY 672 ACCAACCTCTGGAGCAGGAGCGGAGAGGTTGAAAAAGCTCCTTGAACAAGAAAAAGCT 731
Db :::::::::::::::::::::
QY 941 ---MetGlnGlnAsnIleGlnGluLeuGluGlnLeuGluGluGluSer 957
QY 732 TACCAAGCCGCAAGAAAGGAAACCGCT-----AAGCGGCTCAACAAACTTCGA 782
Db :::::::::::::::::::::
QY 958 AlaArgGlnLysLeuGlnLeuGluLysValThrThrGluAlaLysLeuLysLysLeuGlu 977
QY 783 GATGAGCTTGTGAAGCTCAAGTCTTCGCCCTCATGTTGTGGACGAGAGCGCAGATGCAC 842
Db :::::::::::::::::::::
QY 978 GluAspValIleValLeuGluAspGlnAsnLeuLysLeuAlaLysGluLysLys 995
QY 843 ATCGAGCAACTGGGCTCGAGAGTCAGAAAGTCCAGGACCTCAGAACCTGAGGGAG 902
Db :::::::::::::::::::::
QY 996 ---LeuLeuGluAspArgMetSerGluPheThrThrAsnLeuThrGlu 1010
QY 903 GAGGAAGAAAACTCAAAGCGGTCACT----- 929
Db :::::::::::::::::::::
QY 1011 GluGluGluLysSerLysSerLeuAlaLysLeuLysAsnLysHisGluAlaMetIleThr 1030

QY 930 -----TACAAATCCAAGGAAGACCCCAAGAGCTGCTCAAGTTAGAA 971
Db 1031 AspLeuGluGluArgLeuArgArgGluLysGlnArgGlnGluLeuGluLysThrArg 1050
QY 972 GTGGACTTCGAACACACAGGCCTCG----- 995
Db 1051 ArgLysLeuGluGlyAspSerSerAspLeuHisAspGlnIleAlaGluLeuGlnAlaGln 1070
QY 996 -----AGTTTTCAGGAGCACCAAGAGATGAACGCCAAATTTGGCG 1037
Db 1071 IleAlaGluLeuLysIleGlnLeuSerLysLysGluGluLeuGlnAlaAlaLeuAla 1090
QY 1038 AATCAAGAATCTCACAAACCGCAACTTCGACTCAAACTGGTTGGCTTATCGCAAAGGATT 1097
Db 1091 ArgValGluGluGluAlaAlaGlnLysAsnMetAlaLeu-----LysLysIle 1106
QY 1098 GAGGAGCTGGAAGAGACCAATAAAGCCTTCAGAAAGGCAGAGGAGCTCCAGGAGCTG 1157
Db 1107 ArgGluLeuGluSerGlnIleThrGluLeuGln-----GluAspLeuGluSerGlu 1123
QY 1158 AGAGAGAAAATTGCCAAAGGGGAATGTGAAACTCCAGTCTCATGGCGGAAGTGGAGAT 1217
Db 1124 ArgAlaSerArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAla 1143
QY 1218 CTGCGCAAGCGCTG-----CTTGAGATGGAGGGCAAGGATGAAGAGATC---ACG 1265
Db 1144 LeuLysThrGluLeuGluAspThrLeuAspSerThrAlaAlaGlnGlnGluLeuArgSer 1163
QY 1266 AAGACCAGAGCCCGAGTGCCTGAGAGTGAAGAAGAGCTCCAAAGAGGAA---GAACACCCAC 1322
Db 1164 LysArgGluGlnGluValThrValLeuLysLysThrLeuGluAspGluAlaLysThrHis 1183
QY 1323 AGCAAGGAACCTTAGACTAGAAAGTGGAGAAGCTGCAGAAAGAGGATGTCTGAGCTGGAGAAG 1382
Db 1184 GluAlaGlnIleGln-----GluMetArgGlnLysHisSerGlnAlaIleGluGlu 1200
QY 1383 CTGGAGGAAGCGTTACGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAG 1442
Db 1201 LeuAlaGluGlnLeuGluGlnThrLys-----ArgValLysAlaAsnLeuGlu 1216
QY 1443 AAGGAGAAGAACCTAACCAAA-----GACCTGCTGAACGAGCTGAGGTG--- 1487
Db 1217 LysAlaLysGlnAlaLeuGluSerGluArgAlaGluLeuSerAsnGluValLysValLeu 1236
QY 1488 -----GTCAAGAGTCGAGTTAAAGAA 1508
Db 1237 LeuGlnGlyLysGlyAspAlaGluHisLysArgLysLysValAspAlaGlnLeuGlnGlu 1256
QY 1509 CTCGAATGCTCCGAGAGTAGACTGGAGAAGCGCCGAGTTAAGCCTCAAAGATGACCTTACA 1568
Db 1257 LeuGlnValLysPheThrGluGlyGluArgValLysThrGluLeuAlaGluArgValAsn 1276
QY 1569 AAGCTGAAGTCTTCACTGTGTGATGCTGCTGGATGAGAGGAAAAATATGATGGAGAAAAATA 1628
Db 1277 LysLeuGlnVal-----GluLeuAspAsnValThrGlyLeuLeu 1289
QY 1629 AAGCAAGAAGAGAGGAAAGTGGATGGGTTGAATAAAAACTTTAAGGTGGAGCAGGGAAAA 1688
Db 1290 AsnGlnSerAspSerLysSerIleLysLeuAlaLysAspPheSerAlaLeuGluSerGln 1309
QY 1689 GTCATGGATGTGACGGAAGAGCTAATCGAGGAAAGCAAGAGCTTTTAAACTCAAATCT 1748
Db 1310 LeuGlnAspThrGlnGluLeuGlnGluThrArgLeuLysLeuSerPheSerThr 1329
QY 1749 GAAATGGAGGAAAAGGATACAGTCTGTGACAAAGGAGAGGATGAGCTGATGGGTAAACTG 1808
Db 1330 LysLeuLysGlnThrGlu-----AspGluLysAsnAlaLeuLysGluGlnLeu 1345
QY 1809 AGGAGCGAAGAAAGGTCCTGTGAACCTGAGCTGCAGTGPAGACTTACTA----- 1859
Db 1346 GluGluGluGluAlaLysArgAsnLeuGluLysGlnIleSerValLeuGlnGlnGln 1365
QY 1860 -----AAGAAGCGGCTTGAT-----GGCATAGAGGAGGTAGAAAGGGAA 1898

DR PROSITE; PS50245; CAP_GLY_2; 2.
KW Cytoskeleton; Microtubule; Coiled coil; Repeat; Alternative splicing.
FT DOMAIN 78 120 CAP-GLY 1.
FT DOMAIN 143 204 SER-RICH.
FT DOMAIN 232 274 CAP-GLY 2.
FT DOMAIN 304 331 SER-RICH.
FT DOMAIN 350 1342 COILED COIL (POTENTIAL).
FT DOMAIN 1408 1421 CCHC-BOX.
FT VARSPLIC 457 491 Missing (in isoform Short).
FT CONFLICT 1069 1069 D -> E (IN REF. 2).
SQ SEQUENCE 1427 AA; 160989 MW; 0A4F166DD94254E8 CRC64;
Alignment Scores:
Pred. No.: 2.29e-11 Length: 1427
Score: 447.00 Matches: 196
Percent Similarity: 43.62% Conservative: 187
Best Local Similarity: 22.32% Mismatches: 329
Query Match: 5.78% Indels: 166
DB: 1 Gaps: 35
US-10-788-793-1 (1-4364) x REST_HUMAN (1-1427)
QY 174 GAAGATGCAAAAAAGAACAAAGGCCAAATCGAAGGAGGAGGAGATGTCATGGCTTCCGGAAC 233
Db 476 GluLysThrLysAlaAspLysLeuGlnArgGluLeuGluAspThrArgValAlaThrVal 495
QY 234 -----ATCAAAAGGCACCTCAAAACCATCTGGAGAAAAGTCAG 269
Db 496 SerGluLysSerArgIleMetGluLeuGluLysAspLeuAlaLeuArgValGlnGluVal 515
QY 270 AAAAAGACTAAGAAGTCTGTGGAGTTATCCAAGGAG-----GACCTCATCCAGCTCCTG 323
Db 516 AlaGluLeuArgArgLeuGluSerAsnLysProAlaGlyAspValAspMetSerLeu 535
QY 324 AGTATCATGGAAGGGAGTTGCAGGCTCAGAAAGATGTCATCCACATGCTGAGGACAGAG 383
Db 536 SerLeuLeuGln--GluIleSerSerLeuGlnGluLysLeuGluValThrArgThrAsp 554
QY 384 AAAACCAAGCCCGAGGTT-----CTGGAGGCACACTATGGATCTGCAGAACCT----- 431
Db 555 HisGlnArg--GluIleThrSerLeuLysGluHisPheGlyAlaArgGluGluThrHis 573
QY 432 GAGAAAGTGCTTCGGGTCCTGCACCGAGATGCCATCCTTGCTCAAGAGAAGTCCATAGGA 491
Db 574 GlnLysGluIleLysAlaLeuTyr-----ThrAlaThrGluLysLeuSerLys 589
QY 492 GAAGACGTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAAAAGCAGAAGGAG 551
Db 590 GluAsn-----GluSerLeuLysSerLysLeuGluHisAlaAsnLysGluAsnSerAsp 607
QY 552 ACGTACCGCGCATGCTAGAGCAGCTGCTGCTGGCTGAGAAGTGTCAcAGGCGCACCGTG 611
Db 608 ValIleAlaLeuTrpLysSerLysLeuGluThrAlaIleAlaSerHisGlnGlnAlaMet 627
QY 612 TACGAGCTGGAGAACGAGAAGCACAAG-----
Db 628 GluGluLeuLysValSerPheSerLysGlyLeuGlyThrGluThrAlaGluPheAlaGlu 647
QY 639 -----CACACTGACTACATGAACAAGAGCGACGACTTCACC 674
Db 648 LeuLysThrGlnIleGluLysMetArgLeuAspTyrGlnHisGluIleGluAsnLeuGln 667
QY 675 AACCTGCTGGAGCAGGAGCAGAGAGGTTGAAAAAGCTCCTTGAA-----CAAGAAAAA 728
Db 668 AsnGlnGlnAspSerGluArgAlaAlaHisAlaLysGluMetGluAlaLeuArgAlaLys 687
QY 729 GCTTACCAAGCCCGCAAGAAAAGGAAAACGCT-----AAGCGGCTCAAC 773
Db 688 LeuMetLysValIleLysGluLysGluAsnSerLeuGluAlaIleArgSerLysLeuAsp 707
QY 774 AAACCTTCAGATGAG---CTTGTGAAGCTCAAGTCCTTCGCCCTCATGTTGGTGGACGAG 830

Db 708 LysAlaGluAspGlnHisLeuValGluMetGluAsp---ThrLeuAsnLysLeuGlnGlu 726
QY 831 AGGCAGATGCACATCCAGCAAACTGGGCCTG-----CAGAGTCAG 869
Db 727 AlaGluIleLysValLysGluLeuGluValLeuGlnAlaLysCysAsnGluGlnThrLys 746
QY 870 AAAGTCCAGGACCTCACTCAGAAAGCTCAGGAGGAGGAGGAAGAAAA-----CTCAAA 920
Db 747 ValIleAspAsnPheThrSerGlnLeuLysAlaThrGluGluLysLeuLeuAspLeuAsp 766
QY 921 GCGGTCACTTACAAATCCAAGGAAGACCGCCAGAAAGCTCTCAAGTTAGAAGTGGACTTC 980
Db 767 AlaLeuArgLysAlaSerSerGluGlyLysSerGluMetLysLysLeuArgGlnGlnLeu 786
QY 981 GAAACACAAGGCCTCGAGGTTTTTCCCAGGAGCACCAAGAGATGAACGCCAAATTTGGCGAAT 1040
Db 787 GluAlaAlaGluLysGlnIleLysHisLeuGluIleGluLysAsnAlaGluSerSerLys 806
QY 1041 CAAGAATCTCAAAACCGGCAACTTCGA-----CTCAAACTGGTTGGCTTATCGCAA 1091
Db 807 AlaSerSerIleThrArgGluLeuGlnGlyArgGluLeuLysLeuThrAsnLeuGlnGlu 826
QY 1092 AGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAAGGCAGAGGAGCTCCAG 1151
Db 827 AsnLeuSerGluValSerGlnValLysGluThrLeuGluLys-----GluLeuGln 843
QY 1152 GAGCTGAGAGAGAAAAATTGCCAAAGGG-----
Db 844 IleLeuLysGluLysPheAlaGluAlaSerGluGluAlaValSerValGlnArgSerMet 863
QY 1179 ---GAATGTGGAAC-----TCCAGTCTCATGGCGGAA 1208
Db 864 GlnGluThrValAsnLysLeuHisGlnLysGluGluGlnPheAsnMetLeuSerSerAsp 883
QY 1209 GTGGAGAGTCTCGCAAGCGCGTGTGATGGAGGCAAGGATGAAGAGATCACGAAG 1268
Db 884 LeuGluLysLeuArgGluAsnLeuAlaAspMetGluAlaLysPheArgGluLysAspGlu 903
QY 1269 ACCGAGGCCAGTGCCTGGAGCTGAGAAGAAGCTCCAAGAGGAA----- 1313
Db 904 ArgGluGluGlnLeuIleLysAlaLysGluLysLeuGluAsnAspIleAlaGluIleMet 923
QY 1313 -----
Db 924 LysMetSerGlyAspAsnSerSerGlnLeuThrLysMetAsnAspGluLeuArgLeuLys 943
QY 1314 GAACACCACAGCAAGGAACTTAGACTAGAAAGTGGAGAGAGCTGCAGAGAGGATGTCAG 1373
Db 944 GluArgAspValGluGluLeuGlnLeuLysLeuThrLysAlaAsnGluAsnAlaSerPhe 963
QY 1374 CTGGAAGAAGCTGGAGGAAGCGTTACGC-----CGGAGTAACTCGGAATGCACC 1421
Db 964 LeuGlnLysSerIleGluAspMetThrValLysAlaGluGlnSerGlnGlnGluAlaIa 983
QY 1422 CAGCTCCATCTGAACCTGGAGAAGGAGAAACCTAACCAAGACCTGCTGAACGAGCTG 1481
Db 984 LysLysHis-----GluGluGluLysLysGluLeuGluArgLysLeuSerAspLeu 1000
QY 1482 GAGGTGGTCAAGAGTCAGATTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAG----- 1535
Db 1001 Glu-----LysLysMetGluThrSerHisAsnGlnCysGlnGluLeu 1014
QY 1536 AAGGCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAGTCCTTCACGTGTGATGCTG 1595
Db 1015 LysAlaArgTyrGluArgAlaThrSerGluThrLysThrLysHisGluGluIleLeu-- 1033
QY 1596 GTGGATGAGAGGAAAAATATGATGGAGAAAAATAAAGCAAGAAGAGAGAAAGTGGATGG 1655
Db 1034 -----GlnAsnLeuGlnLysThrLeuLeuAspThrGluAspLysLeuLysGly 1049
QY 1656 TTGAATAAAAACTTTAAGGTGGAGCGGGAAGAAATCATGGATGTGACGGAA-----AAG 1709
Db 1050 AlaArgGluGlu-----AsnSerGlyLeuLeuGlnGluLeuGluLeuArgLys 1066

Db 1205 uGlnGluLeuArgGlnGluArgAlaArgLysLeuArgGluGluGlnLeuLeuArgG1 1225
QY 2174 CCAGGAAGCCGAACGCGA-----CACAGGTTTCGGCTGGAGGAG----- 2213
Db 1225 nGluGluGlnGluLeuArgGlnGluArgAspArgLysPheArgGluGluGlnLeuLe 1245
QY 2214 -----GCTAAAGTCGTGATTACAGGCCGCGAGGTGCA 2245
Db 1245 uArgArgGluGluGlnGluLeuArgArgGluArgAspArgLysPheArgGluGluG1 1265
QY 2246 GGCTCTCAAGGAGAGATCCACGAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTCCA 2305
Db 1265 nLeuLeuGlnGluArgGluGluGlnGluArgLeuArgGlnGluArgAlaArgLysLeuAr 1285
QY 2306 AGTCGACTATTTCGCTCT-----CAGCAAAGATTTATGGAAGAAGA 2347
Db 1285 gGluGluGluGlnGlnLeuLeuPheGluGluGlnGluGlnArgLeuArgGlnGluAr 1305
QY 2348 AACTAAG-----AACAAGAACATGGGGAGGAGGTCTCAATCTGACCAAGGA 2395
Db 1305 gAspArgArgTyrArgAlaGluGluGlnPheAlaArgGlu----- 1318
QY 2396 GCTAGAGCTTTCCAAAGCGCTACAGCCGAGCTCTCAGCCGAGTGGGAACGGCCGAAGGAT 2455
Db 1319 ---GluLysSerArgArgGluGluArgGluLeuArgGlnGluGluGlnArgArgAr 1337
QY 2456 GGTGACGTGCCTGTGGCCTCCACTGGGTGTCAGACCGAGCGGTGTGCGGGGATGCTGC 2515
Db 1337 gArgGlu-----Ar 1340
QY 2516 GGAGGAGGAGACCCCGCTGTGTTTCATTGCAAAATCCTCCAGGAGGAAAAATCACATCAT 2575
Db 1340 gGluArgLysPheArgGluGluGlnLeuArgArgGln---GlnGluGluGluGlnArgAr 1359
QY 2576 GAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCATGGAACGGTCTCGTCTCGAC 2633
Db 1359 gArgGlnLeuArgGluArgGlnPheArgGluAspGlnSerArgArgGlnValleuGlu 1378

RESULT 12
MYHB_RABIT
ID MYHB_RABIT STANDARD; PRT; 1972 AA.
AC P35748;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, smooth muscle isoform (SMMHC).
GN MYH11.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92073350; PubMed=1961735;
RA Babi J P., Kelly C., Periasamy M.;
RT "Characterization of a mammalian smooth muscle myosin heavy-chain
gene: complete nucleotide and protein coding sequence and analysis of
the 5' end of the gene";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10676-10680(1991).
CC -|- FUNCTION: Muscle contraction.
CC -|- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
and 2 regulatory light chain subunits (MLC-2).
CC -|- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -|- DOMAIN: The rodlike tail sequence is highly repetitive, showing
cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
characteristic for alpha-helical coiled coils.
CC -|- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be
split further into 2 globular subfragments (S1) and 1 rod-shaped
subfragment (S2).
CC -|- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -|- SIMILARITY: Contains 1 IQ domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M77812; AAA31395.1; -.
DR HSSP; P10587; 1BR2.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; myosin_N.
DR InterPro; IPR002928; myosin_tail.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Multigene family.
FT DOMAIN 1 784 MYOSIN HEAD-LIKE.
FT DOMAIN 785 807 IQ.
FT DOMAIN 844 1934 COILED COIL (POTENTIAL).
FT DOMAIN 1935 1972 CARBOXYL-TERMINAL.
FT NP_BIND 178 185 ATP (POTENTIAL).
FT DOMAIN 661 683 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 763 777 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 129 129 METHYLATION (TRI-) (POTENTIAL).
SQ SEQUENCE 1972 AA; 227318 MW; 2061A24288D6A4C CRC64;

Alignment Scores:

Pred. No.:	2,76e-11	Length:	1972
Score:	445.00	Matches:	248
Percent Similarity:	35.79%	Conservative:	164
Best Local Similarity:	21.55%	Mismatches:	333
Query Match:	5.76%	Indels:	406
DB:	1	Gaps:	44

US-10-788-793-1 (1-4364) x MYHB_RABIT (1-1972)

QY 192 AAGCCCAATCGGAAGGAGGAGGATGTCATGGCTTCC-----GGAACATATCAAA 239
Db 846 GlnValThrArgGlnGluGluMetGlnAlaLysGluAspGluLeuGlnLysIleLys 865
QY 240 ---AGGCACCTCAAAACCATCTGGAGAAAGTGAAGAAAAAGACTAAGAGTCTGTGGAGTTA 296
Db 866 GluArgGlnGlnLysAlaGluSerGluLeuGlnGluLeuGlnLysHisThrGlnLeu 885
QY 297 TCC----- 299
Db 886 SerGluGluLysAsnLeuLeuGlnGluGlnLeuGlnAlaGluThrGluLeuTyrAlaGlu 905
QY 300 -----AAGGAGGACCTCATCCAGCTCCTGAGT 326
Db 906 AlaGluGluMetArgValArgLeuAlaLysLysGlnGluLeuGluGluIleLeuHis 925
QY 327 ATCATGGAAGGGAGTTGTCAGGCTCGAAGAGATGTATCCACATGTGTGAGGACAGAGAA 386
Db 926 GluMetGluAlaArgLeuGluGluGluGluAspArgGlyGlnGlnLeuGlnAlaGluArg 945
QY 387 ACCAAG-----CCCGAGGTTCTGGAGGCACACTATGGATCTGCAGAACCTGAGAAAGTG 440
Db 946 LysLysMetAlaGlnGlnMetLeuAspLeuGluGlnLeuGluGluGluAlaAla 965
QY 441 CTTCGGGTCTCTGCACCGAGATGCCATCCTTCTGCTCAAGAG-----AAGTCCATAGGAGAA 494

Db 966 ArgGlnLysLeuGlnLeuGluLysValThrAlaGluAlaLysIleLysLysLeuGluAsp 985
Qy 495 GACGTCATATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAAAAAGCAGAAAGGAGACG 554
Db 986 AspIleLeuValMetAsp-----AspGlnAsnAsnLysLeuSerLysGlu--- 1000
Qy 555 TACCGCCGATGCTAGAGCAGCTGCTG-----CTGGCTGAGAAAGTGT 596
Db 1001 ---ArgLysLeuLeuGluGluArgIleSerAspLeuThrThrAsnLeuAlaGlu----- 1017
Qy 597 CACAGGCGCACCGTGTACGAGCTGGAGAACGAGAACACAGCAC---ACTGACTACATG 653
Db 1018 -----GluGluGluLysAlaLysAsnLeuThrLysLeuLys 1029
Qy 654 AACAAAGCGACGACTTCACCAACCTGCTGGAG----- 686
Db 1030 AsnLysHisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSer 1049
Qy 687 ---CAGGACGCGAGAGGTTGAAAAAGCTCTTGAACAGAA----- 725
Db 1050 ArgGlnGluLeuGluLysLeuLysArgLysMetAspGlyGluAlaSerAspLeuHisGlu 1069
Qy 726 -----AAAGCTTACCAGACCCCGCAAAGAA 749
Db 1070 GlnIleAlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGlu 1089
Qy 750 AAGGAAAAACGCTAAGCGGCTCAACAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCCTTC 809
Db 1090 GluGluLeuGlnAlaAlaLeuAlaArgLeuGluAspGlu---ThrSerGlnLysAsn 1108
Qy 810 GCCCTCATGTTGGTGGACGAGAGGCAGATGCACATCGAGCAACTG-----GGCCTGCAG 863
Db 1109 AlaLeuLysLysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAsp 1128
Qy 864 AGTCAG-----AAAGTCCAGGACCTCACTCAGAACCTGAGGGAGGAGGAA 908
Db 1129 SerGluArgAlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeu 1148
Qy 909 GAAAAACTCAAAGCG-----GTCACTTACAAATCCAAAGGAAGAC 947
Db 1149 GluAlaLeuLysThrGluLeuGluAspThrLeuAspThrAlaThrGlnGlnGluLeu 1168
Qy 948 CGCCAGAG-----CTGCTCAAGTTAGAACTGGACTTCGAA----- 983
Db 1169 ArgAlaLysArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArg 1188
Qy 984 ---CACAAAGGCC-----TCGAGGTTTTCCCGAGGACGACGAAGAGATGAACGCCAAATTG 1034
Db 1189 SerHisGluAlaGlnValGlnGluMetArgGlnLysHisThrGlnValValGluGluLeu 1208
Qy 1035 CGGAATCAA-----LysLeuGluValGlnLeuGlnGluLeuGln 1265
Db 1209 ThrGluGlnLeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysThrLysGlnThr 1228
Qy 1044 ---GAATCTCACAAACCGCAACTTCGACTCAAACCTGGTTGGCTTATCGCAAAGGATTGAG 1100
Db 1229 LeuGluLysGluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGln 1248
Qy 1101 GAGCTGGAAGAGACCATAAAAAGCCTTCAGAAGGCGCAGAGGAAGAGCTCCAGGAGCTCAGA 1160
Db 1249 GluValGluHisLysLysLys-----LysLeuGluValGlnLeuGlnGluLeuGln 1265
Qy 1161 GAGAAAATTGCCAAAGGGGAATGTGGAAACTCC-----AGTCTC 1199
Db 1266 SerLysCysSerAspGlyGluArgAlaArgAlaGluLeuAsnAspLysValHisLysLeu 1285
Qy 1200 ATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTGTCTGAGATGGAGGCAAG----- 1250
Db 1286 GlnAsnGluValGluSerValThrGlyMetLeuSerGluAlaGluGlyLysAlaIleLys 1305
Qy 1251 ---GATGAAGAGATCACGAAGACCGAGGCCAGTGCCTGGGAGCTGAAGAAGAGCTCCAA 1307
Db 1306 LeuAlaLysGluValAlaSerLeuGlySerGlnLeuGlnAspThrGlnGluLeuGln 1325

Qy 1308 GAGGAA----- 1313
Db 1326 GluGluThrArgGlnLysLeuAsnValSerThrLysLeuArgGlnLeuGluAspGluArg 1345
Qy 1314 -----GAACACCAC 1322
Db 1346 AsnSerLeuGlnGluGlnLeuAspGluGluMetGluAlaLysGlnAsnLeuGluArgHis 1365
Qy 1323 AGCAAGGAACCTTAGACTAGAAAGTG-----GAGAAAGCTGCAGAGAGGATGTCT 1370
Db 1366 IleSerThrLeuAsnIleGlnLeuSerAspSerLysLysLysLeuGlnAspPheAlaSer 1385
Qy 1371 GAGCTGGAGAAGCTGGAGGAAGCGTTTCAGCCGAGTAAGTCGGAATGCACCCAGCTC--- 1427
Db 1386 ThrValGluSerLeuGluGluGlyLysLysArgPheGlnLysGluIleGluSerLeuThr 1405
Qy 1428 -----CATCTGAACCTGGAGAAGGAGAAAGAAACCTAACCC 1460
Db 1406 GlnGlnTyrGluLysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeu 1425
Qy 1461 AAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCC 1520
Db 1426 GlnGlnGluLeuAspAspLeuValValAspLeuAspAsnGlnArgGlnLeu----- 1442
Qy 1521 GAGAGTAGACTGGAGAAGCCCGAGTTAAGCCTCAAAGATGACCTTACAAAAGCTGAAGTCC 1580
Db 1443 ValSerAsnLeuGluLysLysGlnLysLysPheAspGlnLeuLeuAlaGluGluLysAsn 1462
Qy 1581 TTCACCTGTGATGCTGCTGGATGAGAGGAAAAAATATGATGGAGAAAAATAAAGCAAGAGAG 1640
Db 1463 IleSerSerLysTyrAlaAspGluArgAspArgAlaGluAlaGluAlaArgGluLysGlu 1482
Qy 1641 AGGAAA-----GTG 1649
Db 1483 ThrLysAlaLeuSerSerLeuAlaArgAlaLeuGluGluAlaLeuGluAlaLysGluGluLeu 1502
Qy 1650 GATGGTTGAATAAAAACTTTAAGGTGGAGCAGGGGAAAAAGTCATG-----CAT 1697
Db 1503 GluArgThrAsnLysMetLeuLysAlaGluMetGluAspLeuValSerSerLysAspAsp 1522
Qy 1698 GTGACGGAAGAGCTAATCGAGGAACAAGAAAGCTTTTAAACTCAAACTCTGAAATGGAG 1757
Db 1523 ValGlyLysAsnValHisGluLeuGluLysSerLysArgAlaLeuGluThrGlnMetGlu 1542
Qy 1758 GAAAGGAGTACAGTCTGACAAAGGAGAGGGATGAGCTGATG-----GGTAAA 1805
Db 1543 GluMetLysThrGlnLeuGluGluLeuGluAspGluLeuGlnAlaThrGluAspAlaLys 1562
Qy 1806 CTGAGGAGCGAA----- 1817
Db 1563 LeuArgLeuGluValAsnMetGlnAlaLeuLysValGlnPheGluArgAspLeuGlnAla 1582
Qy 1818 -----GAAGAAAAGTCTCTGTGAACCTGAGCTGCAGTGTAGAC----- 1853
Db 1583 ArgAspGluGlnAsnGluGluLysArgArgGlnLeuGlnArgGlnLeuHisGluTyrGlu 1602
Qy 1854 -----TTACTAAAGAACGGGCTT 1871
Db 1603 ThrGluLeuGluAspGluArgLysGlnArgAlaLeuAlaAlaLysLysLysLeu 1622
Qy 1872 GATGCG----- 1877
Db 1623 GluGlyAspLeuLysAspLeuGluLeuGlnAlaAspSerAlaIleLysGlyArgGluGlu 1642
Qy 1878 -----ATAGAGGAGGTAGAAAGGGAAATAAAC 1904
Db 1643 AlaIleLysGlnLeuLeuLysLeuGlnAlaGlnMetLysAspPheGlnArgGluLeuGlu 1662
Qy 1905 CGAGGTAGGTGCTCAAGGGGTCTGAGTTCCACCTGCCCGGAAGACAAATAAGATCAGAGAA 1964
Db 1663 AspAlaArgAlaSerArgAspGluIlePheAlaThrAlaLysGluAsnGluLysLysAla 1682

QY 1965 CTAACGCTTGAA-----ATCGAGAGACTG 1988
Db 1683 LysSerLeuGluAlaAspLeuMetGlnLeuGlnGluAspLeuAlaAlaGluArgAla 1702
QY 1989 AAGAAACGG-----CTCCAGCAGTTGGAGGTGGTGAG----- 2021
Db 1703 ArgLysGlnAlaAspLeuGluLysGluGluLeuAlaGluLeuAlaSerSerLeuSer 1722
QY 2022 -----GGGGACTTGATGAAGACCGAGCGACGAATATGACCAGTTGGAG 2063
Db 1723 GlyArgAsnAlaLeuGlnAspGluLysArgArgLeuGluAlaArgIleAlaGlnLeuGlu 1742
QY 2064 CAGAAGTTCAGAACCGAGCAG-----GATAAGGCCAACTTCCTC 2102
Db 1743 GluGluLeuGluGluGlnGlnGlyAsnMetGluAlaMetSerAspArgValArgLysAla 1762
QY 2103 TCCCAGCAGCTCGAGGAATCAAAACACCAATGGCCCAAGCACAAGCC---ATACAGAAA 2159
Db 1763 ThrGlnGlnAlaGluGlnLeuSerAsnGluLeuAlaThrGluArgSerThrAlaGlnLys 1782
QY 2160 GGGGAGGCCGTGAGCCAGGAGCCGAA----- 2186
Db 1783 AsnGluSerAlaArgGlnLeuGluArgGlnAsnLysGluLeuLysSerLysLeuGln 1802
QY 2187 -----CTGGACACAGGTTTCGG----- 2204
Db 1803 GluMetGluGlyAlaValLysSerLysPheLysSerThrIleAlaAlaLeuGluAlaLys 1822
QY 2205 -----CTGGAG-----GAGCTAAAGTCGTGATTACAGGCCCGAGTGCAG 2246
Db 1823 IleAlaGlnLeuGluGlnValGluGlnGluAlaArgGluLysGlnAlaAlaLys 1842
QY 2247 GCTCTCAAGGAG-----AAGATCCACGAGCTGATG----- 2276
Db 1843 AlaLeuLysGlnArgAspLysLysLeuLysGluMetLeuLeuGlnValGluAspGluArg 1862
QY 2276 ----- 2276
Db 1863 LysMetAlaGluGlnTyrLysGluGlnAlaGluLysGlyAsnAlaLysValLysGlnLeu 1882
QY 2277 -----AACAAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTATTCGGTC 2321
Db 1883 LysArgGlnLeuGluGluAlaGluGluSerGlnArgIleAsnAlaAsnArgArgLys 1902
QY 2322 CTTCAGCAAAGATTATGGAAGAAGAACTAAGACAAAGAACATGGGGAGGGAGTCTCCTC 2381
Db 1903 LeuGlnArgGluLeuAspGluAlaThrGluSerAsnGluAlaMetGlyArgGluValAsn 1922
QY 2382 AATCTGACCAAGGAGCTAGAGCTTTCCAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGG 2441
Db 1923 AlaLeuLysSerLysLeuArgArgGlyAsnGluThrSerPheValProThrArgArgSer 1942
QY 2442 AACGGCCGAAGGATGTTGGACGTGCCTGTGGCTCCACTGGGGTGCAGACCGAGCGGTG 2501
Db 1943 GlyGlyArgArgValIleGlu----- 1949
QY 2502 TCGGGGATGCTCGGAGGAGGAGACCCCGGCT 2534
Db 1950 AsnAlaAspGlySerGluGluValAspAla 1960

RESULT 13
MYSN_DROME

ID MYSN_DROME STANDARD; PRT; 2017 AA.
AC Q99323;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myosin heavy chain, non-muscle (Zipper protein) (Myosin II).
GN ZIP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RC TISSUE=Embryo;
RX MEDLINE=90349606; PubMed=2117279;
RA Ketchum A.S., Stewart C.T., Stewart M., Kiehart D.P.;
RT "Complete sequence of the Drosophila nonmuscle myosin heavy-chain
transcript: conserved sequences in the myosin tail and differential
splicing in the 5' untranslated sequence.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:6316-6320(1990).
CC -!- FUNCTION: Nonmuscle myosin appears to be responsible for
cellularization. Required for morphogenesis and cytokinesis.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q99323-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q99323-2; Sequence=VSP_003342;
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M35012; AAA28713.1; -;
DR PIR; A36014; A36014.
DR HSSP; P10587; 1BR2.
DR FlyBase; FBgn0005634; zip.
DR GO; GO:0005856; C:cytoskeleton; NAS.
DR GO; GO:0005860; C:non-muscle myosin; NAS.
DR GO; GO:0030018; C:Z disc; IDA.
DR GO; GO:0046663; P:leading edge cell differentiation; IMP.
DR GO; GO:0006936; P:muscle contraction; IMP.
DR GO; GO:0045214; P:sarcomere organization; IMP.
DR GO; GO:0007395; P:spreading of leading edge cells; IMP.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Alternative splicing; Coiled coil; Actin-binding;
KW ATP-binding; Calmodulin-binding.
FT DOMAIN 1 829 MYOSIN HEAD-LIKE.
FT DOMAIN 830 859 IQ.
FT DOMAIN 886 2017 COILED COIL (POTENTIAL).
FT NP_BIND 225 232 ATP.
FT DOMAIN 250 260 25 kDa/50 kDa JUNCTION.
FT DOMAIN 682 694 50 kDa/20 kDa JUNCTION.
FT DOMAIN 705 727 ACTIN-BINDING.
FT DOMAIN 742 758 REACTIVE SULFHYDRYL/ACTIN-BINDING.
FT DOMAIN 1303 2017 LIGHT MEROMYOSIN (LMM).
FT DOMAIN 1303 1970 ALPHA-HELICAL TAILPIECE (LMM).
FT DOMAIN 1971 2017 GLOBULAR TAILPIECE.
FT VARSPLOC 1 45 Missing (in isoform Short).
FT /FTId=VSP_003342.
SQ SEQUENCE 2017 AA; 232016 MW; 73E3CB02BA8F2528 CRC64;

Alignment Scores:

Pred. No.: 3.32e-11 Length: 2017
Score: 443.00 Matches: 271

Percent Similarity: 38.65%		Conservative: 231	
Best Local Similarity: 20.86%		Mismatches: 483	
Query Match: 5.73%		Indels: 314	
DB: 1		Gaps: 51	
US-10-788-793-1 (1-4364) x MYSN_DROME (1-2017)			
QY	215	TGTCATGGCTTC-----CGAACTATCAAAAGGCACCTCAAAACCATC-----	256
Db	841	CysArgGlyPheLeuAlaArgArgAsnTyrGlnLysArgLeuGlnGlnLeuAsnAlaIle	860
QY	257	-----TGGAGA 262	
Db	861	ArgIleIleGlnArgAsnCysAlaAlaTyrLeuLysLeuArgAsnTrpGlnTrpTrpArg	880
QY	263	AAGTGAGAAAAAGACTAAGAAGTCTGTGGAGTTATCCAAAG-----GAGGACCTCATCCA	316
Db	881	-LeuTyrThrLysValLysProLeuLeuGluValThrLysGlnGluLysLeuValG1	900
QY	317	GCTCCTGAGTATCATGGAAGGGGAGTTG---CAGGCTCGAGAAGATGTCTCCACATGCT	373
Db	900	n-----LysGluAspGluLeuLysGlnValArgGlu-----LysLe 912	
QY	374	GAGGACAGAGAAAAACCAAGCCCGAGGTCTGTGGAGGCACACTATGGATCTGCAGAACCTGA	433
Db	912	uAspThrLeuAlaLysAsnThrGlnGluTyrGluArgLysTyrGlnGlnAlaLeuValG1	932
QY	434	GAAG-----GTGCTTCGGGTCTCGCACCGAGATGCCATCCTTGTCTCAAGAGAAGTCCAT	487
Db	932	uLysThrThrLeuAlaGluGlnLeuGlnAlaGluIleGluLeuCysAlaGlu-----	949
QY	488	AGGAGAAGACGCTCTATGAGAAACCTATCTCAGAGCTGCAGACACTGGAGGAAAAAGCAGAA	547
Db	950	-AlaGluGluSerArgSerArgLeuMetAlaArgLysGlnGluLeuGluAspMetMetG1	969
QY	548	GGAGACGTACCGCGCATGCTAGAGCAGCTGCTGTGGCTGAGAAGTGTCTCAGCGGCAC	607
Db	969	nGluLeuGluThrArgIleGluGluGlu-----GluGluAr 981	
QY	608	CGTGTAAGAGCTGGAGAACGAGAGAACCAAGCACACTGACTACATGAACAAGAGCGACGA	667
Db	981	gValLeuAlaLeuGlyGlyGluLysLysLys-----	991
QY	668	CTTCACCAACCTGCTGGAGCAGGAGCGAGAGAGGTTGAAAAAGCTCCTTGAACAAGAAAA	727
Db	992	-----LeuGluLeuAsnIleGlnAspLeuGluGlnLeuGluGluGluG1 1007	
QY	728	AGCTTACCAAGCCCGCAAGAAAGAAACGCTAAG-----CGGCTCAACAAACT 778	
Db	1007	uAlaAlaArgGlnLysLeuGlnLeuGluLysValGlnLeuAspAlaLysIleLysLysTy 1027	
QY	779	TCGAGATGAGCTTGTGAAGCTCAAGTCTCCTCGCCCTCATGTTGGTGGACGAGAGGCAGAT	838
Db	1027	rGluGluAspLeuAla-----LeuThrAspAspGlnAsnGlnLy 1040	
QY	839	GCACATCGAGCAACTGGGCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAAGCTGAG	898
Db	1040	sLeuLeuLysGluLysLysLeuLeuGluGluArgAlaAsnAspLeuSerGlnThrLeuAl	1060
QY	899	GGAGGAGGAAGAAAAACTCAAGCGGTCACT-----	929
Db	1060	aGluGluGluGluLysAlaLysHisLeuAlaLysLeuLysAlaLysHisGluAlaThrIl	1080
QY	930	-----TACAAATCCAAGGAAGACCGCCAG-----	953
Db	1080	eThrGluLeuGluGluArgLeuHisLysAspGlnGlnArgGlnGlnSerArgSe 1100	
QY	954	----AAGCTGCTCAAGTTAGAAGTG---GACTTCGAACACAAGGCCTCGAGGTTTCCCA	1006
Db	1100	rLysArgLysIleGluThrGluValAlaAspLeuLysGluGlnLeuAsnGluArgVa 1120	
QY	1007	GGAGCACGAAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCACACCGCGCAACTTCG	1066

Db	1120	lGlnValAspGluMetGlnAlaGlnLeuAlaLysArgGluGluGluLeuThrGlnThrLe	1140
QY	1067	ACTCAAACTGGTTGGCTTATCGAAAGGATTGAGGAGCTGGAAGAGACCAATAAAGCCT	1126
Db	1140	uLeu-----ArgIleAspGluSerAlaThrLysAlaThrAl	1153
QY	1127	TCAGAAGGCAGAGGAAGAGCTCCAG-----GAGCTGAGAGAAAAATT-----	1169
Db	1153	aGlnLysAlaGlnArgGluLeuGluSerGlnLeuAlaGluIleGlnGluAspLeuGluAl	1173
QY	1170	-----GCCAAAGGGGAATGTGAAACTCCAGTCTCATGGCGGAAGTGGA	1213
Db	1173	aGluLysAlaAlaArgAlaLysAlaGluLysValArgArgAspLeuSerGluGluLeuG1	1193
QY	1214	GAGTCTCGCAAGCGCTGCTTGAG-----ATGAGGGCAAGGATGAAGAGAT	1261
Db	1193	uAlaLeuLysAsnGluLeuLeuAspSerLeuAspThrThrAlaAlaGlnGlnGluLeuAr	1213
QY	1262	CACGAAGACCGAGGCCAGTCCCGGAGCTGAAGAAAGAACTCCAAGAGGAA-----	1313
Db	1213	gSerLysArgGluGlnGluLeuAlaThrLeuLysLysSerLeuGluGluThrValas	1233
QY	1314	-----GAACACCACAGCAAGAACTTAGACTA-----	1340
Db	1233	nHisGluGlyValLeuAlaAspMetArgHisLysHisSerGlnGluLeuAsnSerIleAs	1253
QY	1341	---GAAGTGGAGAGCTGCAGAAAGAGGATGTCTGAGCTGGAGAAAGCTGGAGGAAGCGTT	1396
Db	1253	nAspGlnLeuGluAsnLeuArgLysAlaLysThrValLeuGluLysAlaLysGlyThrLe	1273
QY	1397	CAGCCGAGTAAGTCGAA---TGCACCCAGCTCCAT---CTGAACCTGGAGAAGGAGAA	1450
Db	1273	uGluAlaGluAsnAlaAspLeuAlaThrGluLeuArgSerValAsnSerSerArgGlnG1	1293
QY	1451	GAACCTTAACCAAGACCTGCTGAACGAGCTGGAGGTGTCAAGAGTCGAGTTAAAGAACT	1510
Db	1293	uAsnAspArgArg-----ArgLysGlnAlaGluSerGlnIleAlaGluLe	1308
QY	1511	CGAATGCTCCGAGAGTAGACTGGAGAAGGCCGAGTTAAGCCTCAAAGATGACCTTACAAA	1570
Db	1308	uGlnValLysLeuAlaGluIleGluArgAlaArgSerGluLeuGlnGluLysCysThrLy	1328
QY	1571	GCTGAAGTCCTTCACTGTGATGCTGTGGATGAGAGGAAAAATATATGATGAGAAAAATAA	1630
Db	1328	sLeuGln-----GlnGluAlaGluAsnIleThrAsnGlnLeuG1 1341	
QY	1631	GCAAGAAGAGAGGAAAGTGGATGGTTGAATAAAAACTTAAAGGTGGAGCAGGGAAGT	1690
Db	1341	uGluAlaGluLeuLysAlaSerAlaAlaValLysSerAlaSerAsnMetGluSerGlnLe	1361
QY	1691	CATGATGTGACGGAAAGCTAATCGAGGAAAGCAAGAGCTTTTAAACTCAAATCT--	1748
Db	1361	uThrGluAlaGlnGlnLeuLeuGluGluThrArgGlnLysLeuGlyLeuSerSerLy	1381
QY	1749	-----GAAATGGAGGAAAAAGGAGTACAGTCTGACAAAAGGAGAGGGATGATGGG	1801
Db	1381	sLeuArgGlnIleGluSerGluLysGluAlaLeuGlnGluGlnLeuGluGluAspG1	1401
QY	1802	TAAACTGAGGAGCGAAGAAGAAAGTCTCTGTGAACCTGAGCTGCAGTGTAGACTTACTAA	1861
Db	1401	uAlaLysArgAsnTyrGluArgLysLeuAlaGluValThrThrGlnMetGlnGluIleLy	1421
QY	1862	GAAGCGCTTGATGGCATAGAGGAGGTAGAAAAGGGAATAAAACCGAGGTAGTTCGTGCAA	1921
Db	1421	sLysLysAlaGluGluAspAlaAspLeuAlaLysGluLeuGluGluGlyLys-----	1438
QY	1922	GGGCTCTGAGTTTACCTGCTGCCCGGAAGACAATAAGATCAGAGAACTAACGCTTGAATCGA	1981
Db	1439	-----Ly 1439	
QY	1982	GAGACTGAAGAAACGGCTCCAGCAGTTTCGAGGTGGTGGAGGGGACTTGTATGAAGACCGA	2041
Db	1439	sArgLeuAsnLysAspIleGluAlaLeuGluArgGlnValLysGluLeuIleAlaGlnAs	1459

Db 752 LeuThrGlnGlnIleArgAlaSerGluGluLysLeuLeuAspLeuAlaAlaLeuGlnLys 771
QY 1362 AGGATGTCT-----GAGCTGGAGAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAG 1409
Db 772 AlaAsnSerGluGlyLysLeuGluIleGlnLysLeuSerGluGlnLeuGlnAlaAlaGlu 791
QY 1410 TCGGAATGCACCCAGCTCCATCTGAACTTGAGAACCTGGAGAGGAGAGAACCTTAACCAAAGACCTG 1469
Db 792 LysGlnIleGln-----AsnLeuGluThrGluLysValSerAsnLeuThrLys----- 807
QY 1470 CTGAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGA 1529
Db 808 -----GluLeuGlnGlyLysGluGlnLysLeuLeuAspLeuGluLysAsnLeuSerAla 825
QY 1530 CTGGAGAAGCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAGTCCTTCACTGTG 1589
Db 826 ValAsnGlnValLysAspSerLeuGluLysGluLeuGlnLeuLeuLysGluLysPheThr 845
QY 1590 ATGCTGGTGATGAGAGGAGGAAAATATATGATGGAGAAAATAAAGCAAGAGAGGAGGAAAAGTG 1649
Db 846 SerAlaValAspGlyAlaGluAsnAlaGlnArgAlaMetGlnGluThrIleAsnLysLeu 865
QY 1650 GATGGGTTGAATAAAACTTTAAGGTGGAGCAGGGAAGTAATCATGGATGTGACGGGAAAAG 1709
Db 866 -----AsnGlnLysGluGluGlnPheAlaLeuMetSerSerGluLeuGlu 880
QY 1710 CTAATCGAGGAAAGCAAGAACTTTTAAAACTCAAA---TCTGAAATGGAGGAAAAGGAG 1766
Db 881 GlnLeuLysSerAsnLeuThrValMetGluThrLysLeuLysGluArgGluGluArgGlu 900
QY 1767 TACAGTCTGACA-----AAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGC 1814
Db 901 GlnGlnLeuThrGluAlaLysValLysLeuGluAsnAspIleAlaGluIleMetLysSer- 920
QY 1815 GAAGAAGAAAGGTCCTGTGAACCTGAGCTGCAGTGTAGACTTACTAAAGAAGCGGCTTGAT 1874
Db 921 SerGlyAspSerAlaGlnLeuMetLysMetAsnAspGluLeuArgLeuLysGluArg 940
QY 1875 GGCATAGAGGAGGTAGAAAGGGAATAAACCAGGTAGTTCG-----TGC 1919
Db 941 GlnLeuGluGlnIleGlnLeuGluLeuThrLysAlaAsnGluLysAlaValGlnLeuGln 960
QY 1920 AAGGGGTCTGAGTTACCTGCCCGGAAGACAATAAGATCAGAGAACTAACGCTTGAAATC 1979
Db 961 LysAsnValGluGlnThrAlaGlnLysAlaGluGlnSerGlnGlnGluThrLeuLysThr 980
QY 1980 -----GAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTGGTGGAGGGGACTTGATG 2033
Db 981 HisGlnGluGluLeuLysLysMetGlnAspGlnLeuThrAspMetLysLysGlnMetGlu 1000
QY 2034 AAGACCGAGGACGAATATGACCAGTTGGAGCAGAGTTCAGAACCCGAG----- 2081
Db 1001 ThrSerGlnAsnGlnTyrLysAspLeuGlnAlaLysTyrGluLysGluThrSerGluMet 1020
QY 2081 ----- 2081
Db 1021 IleThrLysHisAspAlaAspIleLysGlyPheLysGlnAsnLeuLeuAspAlaGluGlu 1040
QY 2082 -----CAGGATAAGGCAAAACCTTCTCTCCAGCAGCTCGAGGAAATCAAA 2126
Db 1041 AlaLeuLysAlaAlaGlnLysLysAsnAspGluLeuGluThrGlnAlaGluGluLeuLys 1060
QY 2127 CACCAAATGGCCAAGCACACAAGCCATAGAGAAAGGGGAGGCCGTGAGCCAGGAAGCCGAA 2186
Db 1061 LysGlnAlaGluGlnAlaLysAlaAspLysArgAlaGluGluValLeuGlnThrMetGlu 1080
QY 2187 CTGCGACACAGGTTTCGGCTGGAGGAGGCTAAAAGTCGTGATTACAGGCCCGAGTGCAG 2246
Db 1081 -----LysValThrLysGluLysAsp-----AlaIleHisGln 1091
QY 2247 GCTCTCAAGGAGAAGATCCACGAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTCCAA 2306
Db 1092 GluLysIleGluThrLeuAlaSerLeuGluAsnSerArgGlnThrAsnGluLysLeuGln 1111

QY 2307 GTCGACTATTCGGTCTCTCAGCAAAGATTATG-----GAAGAAGAAAACCTAAGAAC 2357
Db 1112 AsnGluLeuAspMetLeuLysGlnAsnAsnLeuLysAsnGluGluLeuThrLysSer 1131
QY 2358 AAGAACATGGGGAGGAGGTCTCAATCTG---ACCAAGGAGCTAGAGCTTTCCCAAGCGC 2414
Db 1132 Lys-----GluLeuLeuAsnLeuGluAsnLysLysValGluGluLeuLysLys 1147
QY 2415 TACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAAGGATGGTGGACGTGCTGTGGCC 2474
Db 1148 GluPheGluAlaLeuLysLeuAlaAlaGlnLysSerGlnGlnLeuAlaAlaLeuGln 1167
QY 2475 TCCACTGGGTGCAGACCGAGCGGTGTGCGGGGATGCTGCGGAGGAGGAGACCCCGCT 2534
Db 1168 GluGluAsnValLysLeuAlaGluGluLeuGlyArgSerArgAspGluValThrSer--- 1186
QY 2535 GTGTTTCATTTCGCAAAATCCTTCCAGGAGGAAAAATCACATCATGATTAAT----- 2582
Db 1187 -----HisGlnLysLeuGluGluArgSerValLeuAsnAsnGlnLeuLeuGlu 1203
QY 2583 -----CTTCGACAGGTAGGCCTGAAGAAAACCCCATG-----GAACGGTCTCGTCTC 2630
Db 1204 MetLysLysArgGluSerThrLeuLysLysGluIleAspGluGluArgAlaSerLeuGln 1223
QY 2631 GACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTGGATTCTCTGGATG 2690
Db 1224 LysSerIleSerAspThrSerAlaLeuIleThrGlnLysAspGluGluLeuGluLysLeu 1243
QY 2691 AGAAAAGAGAAAACGGTCTTCCACTCCGACAGAGAAAGGCCCCAGGCCAAACAGGGT 2750
Db 1244 ArgAsn----- 1245
QY 2751 GCAGGGCACCCCGGGAGCTGGTCTTAGCACCAAAAGCAGGGCCAGCCCTACACATCCGT 2810
Db 1246 -----GluIleThr 1248
QY 2811 GTGACACCAGATCATGAGAACAGCACTGCCACCCTGGAGATCACAAGCCCCACATCTGAA 2870
Db 1249 ValLeuArgGlyGluAsnAlaSerAlaLysThrLeuGln----- 1261
QY 2871 GAGTTTTTCTCTAGTACCACCGTCATTCTCCTACCTTAGGCAACCCAGAAAACCAAGATAACC 2930
Db 1262 -----SerValValLysThrLeuGluSerAspLysLeuLys----- 1273
QY 2931 ATTATTCCATCACCCAATGTCATGTGCGAAAAGCCCCAAAAGTGCAGATCTCTCTCGGC 2990
Db 1274 -----LeuGluGluLysValLysAsnLeuGluGlnLysLeu--- 1285
QY 2991 CCAGAACGAGCCATGTCCCTGTTCACGATTACTACTATTTCAGAGAGAGAGCCCGGAA 3050
Db 1286 LysAlaLysSerGluGlnProLeuThrValThr-----SerProSer 1299
QY 3051 GGTGGAAGGAGCGCC-----TTTGCCGACAGGCCCTGCATCC----- 3086
Db 1300 GlyAspIleAlaAlaAsnLeuLeuGlnAspGluSerAlaGluAspLysGlnGlnGluIle 1319
QY 3087 -----CCC 3089
Db 1320 AspPheLeuAsnSerValIleValAspLeuGlnArgArgAsnGluGluLeuAsnLeuLys 1339
QY 3090 ATCCAAATCATGCGGTGTCAACATCTGCAGCTCCCACCTGAAATCGCTGTCTCTCTGAA 3149
Db 1340 IleGlnArgMetCysGluAlaAlaLeuAsnGlyAsnGluGluGluThrIleAsnTyrAsp 1359
QY 3150 TCTCAGGAAGTGCCTATGGGAAGGACTATCTCAAAGTCACCCCGGAAAAACAAACTGTT 3209
Db 1360 SerGlu-Glu-----GluGlyLeuSerLysThrProArg-----LeuPh 1373
QY 3210 CCAGCCCCCGTGGGA-----AGTACAACTCC 3236
Db 1373 eCysAspIleCysGlyCysPheAspLeuHisAspThrGluAspCysProThrGlnAlaGln 1393

QY 689 GGAGCGAGAG-----AGGTTGAAAAAGACTCCTTGAACAAGAAAAAGC 730
Db 1112 lValArgGluLeuGlnAlaGlnIleAlaGluLeuGlnGluAspPheGluSerGluLysAl 1132
QY 731 TTACCAAGCCCGCAAAGAAAGGAAAAAC-----GCTAAGCGGCTCAACAACACTTCG 781
Db 1132 aSerArgAsnLysAlaGluLysGlnLysArgAspLeuSerGluLeuGluAlaLeuLy 1152
QY 782 AGATGAGCTT----- 791
Db 1152 sThrGluLeuGluAspThrLeuAspThrThrAlaAlaGlnGlnGluLeuArgThrLysAr 1172
QY 792 -----GTGAAGCTCAAGTCCTCGCCCT 814
Db 1172 gGluGlnGluValAlaGluLeuLysLysAlaLeuGluGluGluThrLysSerHisGluAl 1192
QY 815 CATGTTGGTGACGAGAGGAGATGCACATCGAGCAACTGGGCCTGCAGAGTCAGAAAGT 874
Db 1192 aGlnIleGlnAspMetArgGlnArgHisAlaThrAlaLeu----- 1205
QY 875 CCAGGACCTCACTCAGAAAGCTGAGGGAGGAGGAAGAAAAAAGCTCAAAGCGGTCACTTACAA 934
Db 1206 -GluGluLeuSerGluGlnLeu--GluGlnAlaLysArgPheLysAlaAsnLeuGluLy 1224
QY 935 ATCCAAGGAA-----GACCGCCAGAAAGCTGCTC----- 962
Db 1224 sAsnLysGlnGlyLeuGluThrAspAsnLysGluLeuAlaCysGluValLysValLeuGl 1244
QY 963 -AAGTTAGAAGTGGAAGCTTCCAAACACAAAGGCCTCGAGGTTTCCCAGGAGCACGAAGAGAT 1021
Db 1244 nGlnValLysAlaGluSerGluHisLysLysArgLysLysLeuAspAlaGlnValGlnGluLe 1264
QY 1022 GAACGCCAAATTTGGCGAATCAAGAATCTCACAAACCGGCAACTTCGACTCAAACCTGGTTGG 1081
Db 1264 uHisAlaLysVal-----SerGluGlyAspArgLeuArgValGluLeuAlaGl 1280
QY 1082 CTTATCGCAAAGGATTGAG---GAGCTGGAAGAGACCAATAAAAGCCTTCAGAAGGCAGA 1138
Db 1280 uLysAlaAsnLysLeuGlnAsnGluLeuAspAsnValSerThrLeuLeuGluGluAlaGl 1300
QY 1139 GGAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAAGCTCCAGTCT 1198
Db 1300 uLysLys-----GlyIleLysPheAlaLysAspAlaAlaGlyLeuGluSerGl 1316
QY 1199 CATGGCGGAAGTGAGAGTCTGCCGAAGCGCGTGTGAGATGGAGGGCAAGGATGAA-- 1256
Db 1316 nLeuGlnAspThrGlnGluLeu-----LeuGlnGluGluThrArgGlnLysLe 1332
QY 1257 -GAGATCACGAAGACCGAGGCCAGTCGCCGGAGCTGAAGAAGAAAGCTCCAAGAGGAAGA 1315
Db 1332 uAsnLeuSerSerArgIleArgGlnLeuGluGluArgSerSerLeuGlnGlnGl 1352
QY 1316 ACACCACAGCAAGGAAGCTTAGACTAGAAGTGGAGAAGCTGCAGAAGAGGATGCTCTGAGCT 1375
Db 1352 nGluGluGluGluAlaArgArgSerLeuGluLys-----GlnLe 1366
QY 1376 GGAGAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAA 1435
Db 1366 uGlnAlaLeuGlnAlaGlnLeuThrAspThrLysLysLysValAspAspAspLeuGlyTh 1386
QY 1436 CCTGGAGAAGGAGAAGAACCTTAACCAAGACCTGCTGAACGAGCTGGAGGTGCTCAAGAG 1495
Db 1386 rIleGluAsnLeuGluGluAlaLysLysLysLeuLeuLysAspValGluValLeuSerGl 1406
QY 1496 TCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAGGCCGAGTTAAGCCTCAA 1555
Db 1406 nArgLeuGluGluLysAlaLeuAlaTyAspLysLeuGluLysThrLysThrArgLeuGl 1426
QY 1556 AGATGACCTTACAAAGCTGAAGTCTTCACTGTGATGCTGGTGGATGAGAGGAAAAATAT 1615
Db 1426 nGlnGluLeuAspAspLeu-----LeuValAspLeuAspHisGlnArgGlnIl 1442

QY 1616 GATGGAGAAAAATAAGCAAGAAGAGAGGAAAGTGGATGGGTTG-----AATAAAAA 1666
Db 1442 eValSerAsnLeuGluLysLysGlnLysLysPheAspGlnLeuLeuAlaGluGluLysAs 1462
QY 1667 CTTTAAAGGTGGAGCAGGGAANAAGTCATGGATGTGACCGAA---AAGCTAATCAGAGAAAG 1723
Db 1462 nIleSerAlaArgTyAlaGluGluArgAspArgAlaGluAlaArgGluLysGl 1482
QY 1724 CAAGAAGCTTTTAAAAACTCAAATCTGAAATGGAGGAANAAGGAGTACAGTCTGACAAAGGA 1783
Db 1482 uThrLysAlaLeuSerLeuAlaArgAlaLeuGluGlu---AlaLeuGluAlaArgGluGl 1501
QY 1784 GAGGATGAGCTGATGGGTAAACTGAGGAGCGAAGAAAGAAAGGTCTCTGTGAACTGAGCTG 1843
Db 1501 uAlaGluArgGlnAsnLysGlnLeuArgAlaAspMetGlu-----AspLeuMetSe 1518
QY 1844 CAGTGTAGACTTTACTAAAGAAGCGGCTTGATGGCATAGAGGAGGTAGAAAAGGAAATAAA 1903
Db 1518 rSerLysAspAspValGlyLysAsnValHisGluLeuGluLysSerLysArgAlaLeu-- 1537
QY 1904 CCGAGGTAGTCTGTCAAAGGGGTCTGAGTTCACCTGCCCGAAGACAATAAGATCAGAGA 1963
Db 1538 -----GluGlnGlnValGluGl 1543
QY 1964 ACTAACGCTTGAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAG----- 2012
Db 1543 uMetArgThrGlnLeuGluGluLeuGluAspGluLeuGlnAlaThrGluAspAlaLysLe 1563
QY 2013 -----GTGGTGAGGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAGCA 2065
Db 1563 uArgLeuGluValAsnMetGlnAlaMetLysAlaGlnPheGluArgAsp---LeuGlnTh 1582
QY 2066 GAAGTTCAGAACCCAGCAGGATGAAGGCAAACTTCCTCTCCCAGCAGCTCGAGGAATCAA 2125
Db 1582 rArgAspGluGlnAsnGluGluLysLysArgLeuLeuIleLysGlnValArgGluLeuGl 1602
QY 2126 ACACCAAATGGCCAAGCACAAAGCCCATAGAGAAAGGGAGGCGGTGAGCCAGGAAGCCGA 2185
Db 1602 uAlaGluLeuGluAspGluArgLys---GlnArgAlaLeuAlaValAlaSerLysLysLy 1621
QY 2186 ACTCGACACAGGTTTCGGCTGGAGGAGGTAAAGTCGTGATTTACAGGCCGAGGTGCA 2245
Db 1621 sMet-----GluIleAspLeuLysAspLeuGluAlaGlnIleGl 1634
QY 2246 GGCTCTCAAGGAGAAGATCCACGAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTCCA 2305
Db 1634 uAlaAlaAsnLysAlaArgAspGluValIleLysGlnLeuArgLysLeuGlnAlaGlnMe 1654
QY 2306 AGTCGACTATTTCGTCCTTCAGCAAAGATTATGGAAGAA----- 2345
Db 1654 tLysAspTyr-----GlnArgGluLeuGluAlaArgAlaSerArgAspGl 1670
QY 2346 -----GAACTAAGAAACAAGAACATGGGGAGGGAGGTCTCT 2380
Db 1670 uIlePheAlaGlnSerLysGluSerGluLysLysLeuLysSerLeuGluAlaGluIleLe 1690
QY 2381 CAATCTGACCAAGGAGCTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGG 2440
Db 1690 uGlnLeuGlnGluLeuAlaSerSerGlu-----ArgAlaArgArgHisAlaGl 1707
QY 2441 GAACGGCCGAAGGATGGTGAGCGTGCCTGTGGCTCCACTGGGGTGCAGACCGGCGGT 2500
Db 1707 uGlnGluArgAspGluLeuAlaAspGluIleAlaAsnSerAlaSerGlyLysSerAlaLe 1727
QY 2501 GTCGGGGATGCTCGGAGGAGGAGACCCCGGCTGTGTTTCATTCGCAAAATCTTCCAGGA 2560
Db 1727 uLeuAspGluLysArgArgLeuGluAlaArgIleAlaGlnLeuGluGluLeuGluGl 1747
QY 2561 GGAAAAATCACATCATGAGTAATCTT 2585
Db 1747 uGluGlnSerAsnMetGluLeuLeu 1755

Search completed: September 7, 2004, 13:48:32
Job time : 275.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
OM nucleic - protein search, using frame_plus_n2p model
Run on: September 7, 2004, 13:25:15 ; Search time 442 Seconds
(without alignments)
6230.414 Million cell updates/sec

Title: US-10-788-793-1
Perfect score: 7731
Sequence: 1 ccactgggtttcttcaggga.....aaaaaaaaaaaaaaaaaaaa 4364

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USFTO_spool_p/US10788793/runat_07092004_134021_10811/app_query.fasta_1.4551
-DB=SPTREMBL_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10788793 @CGN_1_1_717 @runat_07092004_134021_10811 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archheap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	6082	78.7	1212 11	Q8k4t4 rattus norv

2	5696.5	73.7	1213	4	Q7Z7B0	Q7z7b0 homo sapien
3	5456.5	70.6	1177	4	Q86TC3	Q86tc3 homo sapien
4	5349	69.2	1140	4	Q9ULE5	Q9ules5 homo sapien
5	4838	62.6	965	11	Q8JZS5	Q8jzs5 rattus norv
6	4541	58.7	965	4	Q96SK6	Q96sk6 homo sapien
7	3051	39.5	653	4	Q9NVI8	Q9nvi8 homo sapien
8	2537.5	32.8	610	11	Q9CS72	Q9cs72 mus musculu
9	2406.5	31.1	512	4	Q8N8B9	Q8n8b9 homo sapien
10	1926	24.9	893	4	Q8N6Z0	Q8n6z0 homo sapien
11	1828	23.6	764	4	Q8IUM3	Q8ium3 homo sapien
12	1645.5	21.3	752	4	Q13597	Q13597 homo sapien
13	893	11.6	1026	4	Q86V48	Q86v48 homo sapien
14	891	11.5	1046	4	Q8TEH1	Q8teh1 homo sapien
15	839	10.9	1067	11	Q99NG3	Q99ng3 mus musculu
16	833	10.8	1068	11	Q8R4U7	Q8r4u7 mus musculu
17	829	10.7	1051	11	Q9ESV1	Q9esv1 rattus norv
18	532	6.9	201	11	Q9CSL6	Q9csl6 mus musculu
19	497.5	6.4	251	11	Q8BQW1	Q8bgw1 mus musculu
20	479	6.2	2139	5	Q07569	Q07569 entamoeba h
21	475.5	6.2	1305	10	Q9FJ35	Q9fj35 arabidopsis
22	474	6.1	2363	4	Q99968	Q99968 homo sapien
23	473	6.1	2055	5	Q8T5C7	Q8t5c7 plasmodium
24	473	6.1	2055	5	Q8IHP3	Q8ihp3 plasmodium
25	471.5	6.1	2007	13	Q02015	Q02015 gallus gall
26	470.5	6.1	1596	5	Q8IJ44	Q8ij44 plasmodium
27	468.5	6.1	1205	5	P92021	P92021 caenorhabdi
28	467	6.0	1909	5	Q25893	Q25893 plasmodium
29	464.5	6.0	1790	3	Q07380	Q07380 saccharomyc
30	464	6.0	1229	5	Q9NJ22	Q9nj22 aequipecten
31	464	6.0	1243	5	Q9NJ21	Q9nj21 aequipecten
32	464	6.0	1253	5	Q9NJ20	Q9nj20 aequipecten
33	464	6.0	1951	5	Q17042	Q17042 aequipecten
34	463.5	6.0	2003	5	Q19658	Q19658 caenorhabdi
35	463	6.0	1960	11	Q8VDD5	Q8vdd5 mus musculu
36	461.5	6.0	1999	11	Q63731	Q63731 rattus norv
37	459.5	5.9	2003	5	Q22869	Q22869 caenorhabdi
38	457	5.9	1219	5	Q9NJ23	Q9nj23 aequipecten
39	456.5	5.9	2328	5	Q9VY43	Q9vy43 drosophila
40	456	5.9	1972	11	Q8R384	Q8r384 mus musculu
41	453.5	5.9	1992	13	Q04834	Q04834 xenopus lae
42	448.5	5.8	1941	5	Q26079	Q26079 placopecten
43	448.5	5.8	1964	13	Q93522	Q93522 xenopus lae
44	448.5	5.8	2360	5	Q8IR55	Q8ir55 drosophila
45	445.5	5.8	1950	5	Q26080	Q26080 placopecten

ALIGNMENTS

RESULT 1
Q8K4T4
ID Q8K4T4 PRELIMINARY; PRT; 1212 AA.
AC Q8K4T4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Filamin-interacting protein L-FILIP.
GN L-FILIP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar;
RX MEDLINE=22100428; PubMed=12055638;
RA Nagano T., Yoneda T., Hatanaka Y., Kubota C., Murakami F., Sato M.;
RT "Filamin A-interacting protein (FILIP) regulates cortical cell
RT migration out of the ventricular zone.";
RL Nat. Cell Biol. 4:495-501(2002).
DR EMBL; AB055759; BAC00851.1; --
DR InterPro; IPR000533; Tropomyosin.
DR PRINTS; PR00194; TROPOMYOSIN.
SQ SEQUENCE 1212 AA; 137752 MW; A1D5B9C5AF7D4F80 CRC64;

Alignment Scores:

Pred. No.: 2.66e-292 Length: 1212
Score: 6082.00 Matches: 1212
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 78.67% Indels: 0
DB: 11 Gaps: 0

US-10-788-793-1 (1-4364) x Q8K4T4 (1-1212)

QY	75	ATGAGATCAAGAAATCAAGGTGGAGAAAGTTCATCTAACGGGCATGTCTCTGCCCCAAG	134
DB	1	MetArgSerArgAsnGlnGlyGlyGluSerSerAsnGlyHisValSerCysProLys	20
QY	135	TCCTCCATCATCAGCAGTGAATGGTGGTAAGGGCCCCCTCAGAAGATGCAAAAAGAACAAAG	194
DB	21	SerSerIleIleSerSerAspGlyGlyLysGlyProSerGluAspAlaLysLysAsnLys	40
QY	195	GCCAAATCGGAAGGAGGAGGTGTCTATGGCTCCGGAACTATCAAAAGGCACCTCAAAACCA	254
DB	41	AlaAsnArgLysGluGluAspValMetAlaSerGlyThrIleLysArgHisLeuLysPro	60
QY	255	TCTGGAGAAAGTGAGAAAAGACTAAGAACTGTGTGGAGTTATCCAAGAGGACCTCATC	314
DB	61	SerGlyGluSerGluLysLysThrLysLysSerValGluLeuSerLysGluAspLeuIle	80
QY	315	CAGCTCCTGAGTATCATGGAAGGGGAGTTGTCAGGGCTCGAAGAGATGTCTATCCACATGCTG	374
DB	81	GlnLeuLeuSerIleMetGluGlyGluLeuGlnAlaArgGluAspValIleHisMetLeu	100
QY	375	AGGACAGAGAAAACCAAGCCCGAGGTTCTGGAGGCACACTATGGATCTGCAGAAACCTGAG	434
DB	101	ArgThrGluLysThrLysProGluValLeuGluAlaHisTyrGlySerAlaGluProGlu	120
QY	435	AAAGTGTCTCGGGTCCTGCACCGAGATGCCATCCTTGCTCAAGAGAAGTCCATAGGAGAA	494
DB	121	LysValLeuArgValLeuHisArgAspAlaIleLeuAlaGlnGluLysSerIleGlyGlu	140
QY	495	GACGTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAAAAGCAGAGGAGACG	554
DB	141	AspValTyrGluLysProIleSerGluLeuAspArgLeuGluLysGlnLysGluThr	160
QY	555	TACCGCCGCATGCTAGACAGCTGCTGTGGCTGAGAAAGTGTACAGGCGCACCGTGTAC	614
DB	161	TyrArgArgMetLeuGluGlnLeuLeuLeuAlaGluLysCysHisArgArgThrValTyr	180
QY	615	GAGCTGGAGAACGAGAACACACACTGACTACATGAACAAGAGCGACGACTTCACC	674
DB	181	GluLeuGluAsnGluLysHisLysHisThrAspTyrMetAsnLysSerAspPheThr	200
QY	675	AACCTGCTGGAGCAGGACGAGAGGTTGAAAAAGCTCTTGAACAGAAAAAGCTTAC	734
DB	201	AsnLeuLeuGluGlnGluArgGluArgLeuLysLysLeuLeuGluGlnGluLysAlaTyr	220
QY	735	CAAGCCCGCAAGAAAGGAAAAACGCTAAGCGGCTCAACAAACTTCGAGATGAGCTGTG	794
DB	221	GlnAlaArgLysGluLysGluAsnAlaLysArgLeuAsnLysLeuArgAspGluLeuVal	240
QY	795	AAGCTCAAGTCTCTCGCCCTCATGTTGGTGGACGAGAGGCAGATGCACATCGAGCAACTG	854
DB	241	LysLeuLysSerPheAlaLeuMetLeuValAspGluArgGlnMetHisIleGluGlnLeu	260
QY	855	GGCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAGCTGAGGGAGGAGGAAGAAAA	914
DB	261	GlyLeuGlnSerGlnLysValGlnAspLeuThrGlnLysLeuArgGluGluGluLys	280
QY	915	CTCAAGCGGTCACTTACAAATCCAAGGAAGACCGCCAGAGCTGTCTCAAGTTAGAGTG	974
DB	281	LeuLysAlaValThrTyrLysSerLysGluAspArgGlnLysLeuLeuLysLeuGluVal	300
QY	975	GACTTCGAACACAAGGCCTCGAGGTTTCCAGGAGCACGAAGAGATGAACGCCCAATTG	1034

DB	301	AspPheGluHisLysAlaSerArgPheSerGlnGluHisGluGluMetAsnAlaLysLeu	320
QY	1035	GCGAATCAAGAAATCTCACAACCCGGCAACTTCGACTCAAACTGGTTGGCTTATCGCAAAAG	1094
DB	321	AlaAsnGlnGluSerHisAsnArgGlnLeuArgLeuLysLeuValGlyLeuSerGlnArg	340
QY	1095	ATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAAGGCAGAGGAAGAGCTCCAGGAG	1154
DB	341	IleGluGluLeuGluGluThrAsnLysSerLeuGlnLysAlaGluGluGluLeuGlnGlu	360
QY	1155	CTGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAAACTCAGTCTCATGGCGGAAGTGGAG	1214
DB	361	LeuArgGluLysIleAlaLysGlyGluCysGlyAsnSerSerLeuMetAlaGluValGlu	380
QY	1215	AGTCTCGCAAGCGCGTCTTGAGATCGAGGGCAAGGATGAAGAGATCACGAAGACCCGAG	1274
DB	381	SerLeuArgLysArgValLeuGluMetGluGlyLysAspGluGluIleThrLysThrGlu	400
QY	1275	GCCAGTGC CGGAGCTGAAAGAAAGCTCCAAGAGGAAGAACACCAAGCAAGGAACCTT	1334
DB	401	AlaGlnCysArgGluLeuLysLysLysLeuGlnGluGluHisHisSerLysGluLeu	420
QY	1335	AGACTAGAAAGTGGAGAAAGCTGCAGAGAGGATGTCTGAGCTGGAGAAAGCTGGAAAGCG	1394
DB	421	ArgLeuGluValGluLysLeuGlnLysArgMetSerGluLeuGluLysLeuGluGluAla	440
QY	1395	TTCAGCGCGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTTGGAAAGGAGAGAAGAC	1454
DB	441	PheSerArgSerLysSerGluCysThrGlnLeuHisLeuAsnLeuGluLysGluLysAsn	460
QY	1455	CTAACCAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAA	1514
DB	461	LeuThrLysAspLeuLeuAsnGluLeuGluValValLysSerArgValLysGluLeuGlu	480
QY	1515	TGCTCCGAGAGTAGACTGGAGAAAGCCGAGTTAAGCTCAAGATGACCTTACAAAGCTG	1574
DB	481	CysSerGluSerArgLeuGluLysAlaGluLeuSerLeuLysAspAspLeuThrLysLeu	500
QY	1575	AAGTCTTCACTGTGATGCTGGTGGATGAGAGGAAAAATATGATGGAGAAAAATAAGCAA	1634
DB	501	LysSerPheThrValMetLeuValAspGluArgLysAsnMetMetGluLysIleLysGln	520
QY	1635	GAAGAGAGAAAGTGGATGGTGGTTGAATAAAAACTTTAAGGTGGAGCGGGAAGAGTCATG	1694
DB	521	GluGluArgLysValAspGlyLeuAsnLysAsnPheLysValGluGlnGlyLysValMet	540
QY	1695	GATGTGACGGAAAAGCTAATCGAGGAAAAGCAAGAGCTTTTAAAACTCAAAATCTGAAATG	1754
DB	541	AspValThrGluLysLeuIleGluSerLysLysLeuLeuLysLeuLysSerGluMet	560
QY	1755	GAGGAAAAGGAGTACAGTCTGACAAAGGAGAGGGATGAGCTGATGGGTAAACTGAGAGC	1814
DB	561	GluGluLysGluTyrSerLeuThrLysGluArgaspGluLeuMetGlyLysLeuArgSer	580
QY	1815	GAAGAAAGAAAGTCTCTGTGAACCTGAGCTGCAGTGTAGACTTACTAAAGAAAGCGGCTTGAT	1874
DB	581	GluGluGluArgSerCysGluLeuSerCysSerValAspLeuLeuLysLysArgLeuAsp	600
QY	1875	GGCATAGAGGAGGTAGAAAAGGGAATAAAACCGAGTGGTTCGTGCAAGGGGTCTGAGTTC	1934
DB	601	GlyIleGluGluValGluArgGluIleAsnArgGlyArgSerCysLysGlySerGluPhe	620
QY	1935	ACCTGCCCGGAAGACAATAAGATCAGAGAACTAACGCTTGAATTCGAGAGACTGAGAAAA	1994
DB	621	ThrCysProGluAspAsnLysIleArgGluLeuThrLeuGluIleGluArgLeuLysLys	640
QY	1995	CGGCTCCAGCAGTTGGAGGTGGTGGAGGGGACTTGATGAAGACCGAGGACGAATATGAC	2054
DB	641	ArgLeuGlnGlnLeuGluValValGluGlyAspLeuMetLysThrGluAspGluTyrAsp	660
QY	2055	CAGTTGGAGCAGAAAGTTCAGAACCGAGCAGGATAAGGCAAACTTCTCTCTCCAGCAGCTC	2114
DB	661	GlnLeuGluGlnLysPheArgThrGluGlnAspLysAlaAsnPheLeuSerGlnGlnLeu	680

QY 2115 GAGGAAATCAAACACCAATGGCCAAAGCACAAAGCCATAGAGAAAAGGGAGGCCGTGAGC 2174
Db 681 GluGluIleLysHisGlnMetAlaLysHisLysAlaIleGluLysGlyGluAlaValSer 700
QY 2175 CAGGAAGCCGAACCTCGCACACAGGTTTCGGCTGGAGGAGGCTAAAAGTCGTGATTACAG 2234
Db 701 GlnGluAlaGluLeuArgHisArgPheArgLeuGluAlaLysSerArgAspLeuGln 720
QY 2235 GCCGAGGTGCAGGCTCTCAAGGAGAAGATCCACGAGCTGATGAACAAGGAAGACCAGCTG 2294
Db 721 AlaGluValGlnAlaLeuLysGluLysIleHisGluLeuMetAsnLysGluAspGlnLeu 740
QY 2295 TCTCAGCTCCAAGTCGACTATTTCGGTCTCTTCCAGCAAGATTTATGGAAGAAGAACTAAG 2354
Db 741 SerGlnLeuGlnValAspTyrSerValLeuGlnArgPheMetGluGluGluThrLys 760
QY 2355 AACAAAGAACATGGGGAGGAGGTCTCTCAATCTGACCAAGGAGCTAGAGCTTCCAAAGCGC 2414
Db 761 AsnLysAsnMetGlyArgGluValLeuAsnLeuThrLysGluLeuGluLeuSerLysArg 780
QY 2415 TACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAGGATGGTGACGTGCCTGTGGCC 2474
Db 781 TyrSerArgAlaLeuArgProSerGlyAsnGlyArgArgMetValAspValProValAla 800
QY 2475 TCCACTGGGGTGACAGCCGAGCGGTGTGCGGGGATGCTGCGGAGGAGGACCCCGGCT 2534
Db 801 SerThrGlyValGlnThrGluAlaValCysGlyAspAlaAlaGluGluGluThrProAla 820
QY 2535 GTGTTTCATTTCGCAATCTTCCAGGAGGAAATCACATCATGAGTAATCTTCGACAGGTA 2594
Db 821 ValPheIleArgLysSerPheGlnGluGluAsnHisIleMetSerAsnLeuArgGlnVal 840
QY 2595 GGCCTGAAGAAACCCATGGAACGGTCTCGGTCTCGACAGGTATCCCCAGCAGCGAAT 2654
Db 841 GlyLeuLysLysProMetGluArgSerSerValLeuAspArgTyrProProAlaAlaAsn 860
QY 2655 GAGCTCACCATGAGGAAGTCTTGATTCTCTGGATGAGAAAAAGAGAAACGGTCTTCC 2714
Db 861 GluLeuThrMetArgLysSerTrpIleProTrpMetArgLysArgGluAsnGlyProSer 880
QY 2715 ACTCCGAGGAGAAAGSGCCAGGCCCAAACCAAGGTGCAGGGCACCCCGGGAGCTGGTC 2774
Db 881 ThrProGlnGluLysGlyProArgProAsnGlnGlyAlaGlyHisProGlyGluLeuVal 900
QY 2775 CTAGCACCAAAGCAGGGCCAGCCCCTACACATCCGTGTGACACCAGATCATGAGAACAGC 2834
Db 901 LeuAlaProLysGlnGlyGlnProLeuHisIleArgValThrProAspHisGluAsnSer 920
QY 2835 ACTGCCACCCTGGAGATCACAAAGCCCCACATCTGAAGAGTTTTTCTCTAGTACCACCGTC 2894
Db 921 ThrAlaThrLeuGluIleThrSerProThrSerGluGluPhePheSerSerThrThrVal 940
QY 2895 ATTCTTACCTTAGGCAACCAAGAAACCAAGAATAACCATATTTCATCACCCCAATGTTCATG 2954
Db 941 IleProThrLeuGlyAsnGlnLysProArgIleThrIleIleProSerProAsnValMet 960
QY 2955 TCGCAAAAGCCCCAAAGTCAGATCCTACTCTCGGGCCAGAACGAGCCATGTCCCTGTGC 3014
Db 961 SerGlnLysProLysSerAlaAspProThrLeuGlyProGluArgAlaMetSerProVal 980
QY 3015 ACGATTACTACTATTTCAGAGAGAAGAGCCCGGAAGGTGGAAGGAGCGCCTTTCGCGAC 3074
Db 981 ThrIleThrThrIleSerArgGluLysSerProGluGlyGlyArgSerAlaPheAlaAsp 1000
QY 3075 AGGCCTGCATCCCCCATCCAAATCATGACGGTGTCAACATCTGCAGCTCCCACTGAAATC 3134
Db 1001 ArgProAlaSerProIleGlnIleMetThrValSerThrSerAlaAlaProThrGluIle 1020
QY 3135 GCTGTCTCTCCTGAATCTCAGGAAGTGCCTATGGGAAGGACTATCCTCAAAGTCACCCCG 3194
Db 1021 AlaValSerProGluSerGlnGluValProMetGlyArgThrIleLeuLysValThrPro 1040

QY 3195 GAAAAACAAACTGTTCAGCCCCCGTGGGAAAGTACAACTCCAATGCTATATCATCACC 3254
Db 1041 GluLysGlnThrValProAlaProValArgLysTyrAsnSerAsnAlaAsnIleThr 1060
QY 3255 ACGGAAGACAAATAAATTACATTACCTGGGTCTCAGTTTAAAGCGATCTCTGGGCCT 3314
Db 1061 ThrGluAspAsnLysIleHisIleHisLeuGlySerGlnPheLysArgSerProGlyPro 1080
QY 3315 GCCGTGAAGCGTGAGCCCAAGTTATCACCGTCCGGCCTGTCAACGTGACAGCGAGAAG 3374
Db 1081 AlaAlaGluGlyValSerProValIleThrValArgProValAsnValThrAlaGluLys 1100
QY 3375 GAGGTTTCTACAGGCACAGTCTTTCGCTCTCCAGGAACCACTCTCTTCAAGACCCGGT 3434
Db 1101 GluValSerThrGlyThrValLeuArgSerProArgAsnHisLeuSerSerArgProGly 1120
QY 3435 GCTAGCAAAGTGACCAGCACTATAACTATAACCCCGGTCAACAACGTCAACACGAGGA 3494
Db 1121 AlaSerLysValThrSerThrIleThrIleThrProValThrThrSerSerThrArgGly 1140
QY 3495 ACCCAATCAGTGTGAGGACAAAGATGGGTCTATCAGCGGCCTACCCCCACCGCATTCCT 3554
Db 1141 ThrGlnSerValSerGlyGlnAspGlySerSerGlnArgProThrProThrArgIlePro 1160
QY 3555 ATGTCAAAGGTATGAAAGCTGGAAAGCCAGTAGTGGCAGCCTCAGGAGCAGGAAATCTG 3614
Db 1161 MetSerLysGlyMetLysAlaGlyLysProValValAlaAlaSerGlyAlaGlyAsnLeu 1180
QY 3615 ACCAAATTCCAGCCTCGAGCTGAGACTGAGACTCAGTCTATGAAATAGAGCTGAAGAAATCTGCA 3674
Db 1181 ThrLysPheGlnProArgAlaGluThrGlnSerMetLysIleGluLeuLysLysSerAla 1200
QY 3675 GCCAGCAGCACTGCCTCTCTTGGAGGGGGGAAGGGC 3710
Db 1201 AlaSerSerThrAlaSerLeuGlyGlyGlyLysGly 1212
RESULT 2
Q7Z7B0 PRELIMINARY; PRT; 1213 AA.
AC Q7Z7B0;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE L-FILIP.
GN L-FILIP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Nagano T., Sato M.;
RT "Human orthologue of L-FILIP.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB086011; BAC77067.1; -.
SQ SEQUENCE 1213 AA; 138108 MW; EE4329B03516E6AF CRC64;
Alignment Scores:
Pred. No.: 2.8e-273 Length: 1213
Score: 5696.50 Matches: 1134
Percent Similarity: 96.54% Conservative: 37
Best Local Similarity: 93.49% Mismatches: 41
Query Match: 73.68% Indels: 1
DB: 4 Gaps: 1
US-10-788-793-1 (1-4364) x Q7Z7B0 (1-1213)
QY 75 ATGAGATCACGAAATCAAGGTGGAGAAAGTTCAATCTAAACGGGCGATGTCTCTGCCCAAG 134
Db 1 MetArgSerArgAsnGlnGlyGlyGluSerAlaSerAspGlyHisIleSerCysProLys 20
QY 135 TCCTCCATCATCAGCAGTGATGGTGGTAAGGCCCTCAGAAGATGCA---AAAAAGAAC 191
||||||| ::: ||| ||||| ||||| |||||

Db 21 ProSerIleIleGlyAsnAlaGlyGluLysSerLeuSerGluAspAlaLysLysLys 40
QY 192 AAGCCCAATCGGAAGGAGGAGTGTCTATCGCTTCGGAACCTATCAAAAGGCACCTCAA 251
Db 41 LysSerAsnArgLysGluAspValMetAlaSerGlyThrValLysArgHisLeuLys 60
QY 252 CCATCTGGAGAAAGTGAGAAAAAGACTAAGAAGTCTGTGGAGTTATCCAAGGAGGACCTC 311
Db 61 ThrSerGlyGluCysGluArgLysThrLysLysSerLeuGluLeuSerLysGluAspLeu 80
QY 312 ATCCAGCTCCTGAGTATCATGGAAGGGAGTTGCAGGCTCGAGAAGATGTCTATCCACATG 371
Db 81 IleGlnLeuLeuSerIleMetGluGlyGluLeuGlnAlaArgGluAspValIleHisMet 100
QY 372 CTGAGGACAGAGAAAAACCAAGCCCCGAGGTTCTGGAGGCACACTATGGATCTGCAGAACCT 431
Db 101 LeuLysThrGluLysThrLysProGluValLeuGluAlaHisTyrGlySerAlaGluPro 120
QY 432 GAGAAAGTGCTTCGGGTCTCGCACCGAGATGCCATCCTTGCTCAAGAGAGTCCATAGGA 491
Db 121 GluLysValLeuArgValLeuHisArgAspAlaIleLeuAlaGlnGluLysSerIleGly 140
QY 492 GAAGACGCTCTATGAGAAACCTATCTCAGAGCTGCACAGACTGGAGGAAAGACAGAAGGAG 551
Db 141 GluAspValTyrGluLysProIleSerGluLeuAspArgLeuGluGluLysGlnLysGlu 160
QY 552 ACGTACCGCGCATGCTAGAGCAGCTGCTGCTGGCTGAGAAGTGTACAGGCGCACCCGTG 611
Db 161 ThrTyrArgArgMetLeuGluGlnLeuLeuAlaGluLysCysHisArgArgThrVal 180
QY 612 TACGAGCTGGAGAACGAGAAGCACAAAGCACACTGACTACATGAACAAGACGACGACTTC 671
Db 181 TyrGluLeuGluAsnGluLysHisLysHisThrAspTyrMetAsnLysSerAspPhe 200
QY 672 ACCAACCTCTGGAGCAGGACGAGAGAGTGTGAAAAGCTCCTTGAACAAGAAAAAGCT 731
Db 201 ThrAsnLeuLeuGluGlnGluArgGluArgLeuLysLysLeuLeuGluGlnGluLysAla 220
QY 732 TACCAAGCCCGCAAGAAAGGAAAAACGCTAAGCGGCTCAACAACTTCGAGATGAGCTT 791
Db 221 TyrGlnAlaArgLysGluLysGluAsnAlaLysArgLeuAsnLysLeuArgAspGluLeu 240
QY 792 GTGAAGCTCAAGTCTTCGCCCTCATGTTGGTGGACGAGAGGCAGATGCACATCGAGCAA 851
Db 241 ValLysLeuLysSerPheAlaLeuMetLeuValAspGluArgGlnMetHisIleGluGln 260
QY 852 CTGGGCTTCAGAGTCAAGAAAGTCCAGGACCTCACTCAGAAGCTGAGGAGGAGGAAGAA 911
Db 261 LeuGlyLeuGlnSerGlnLysValGlnAspLeuThrGlnLysLeuArgGluGluGlu 280
QY 912 AAACCTCAAGCGGTCACTTACAAATCCAAGGAAGACCGCCAGAAAGCTGCTCAAGTTAGAA 971
Db 281 LysLeuLysAlaIleThrSerLysSerLysGluAspArgGlnLysLeuLeuLysLeuGlu 300
QY 972 GTGGACTTCGAACACAAAGGCTCGAGGTTTCCAGGAGCACGAAGAGATGAACGCCAAA 1031
Db 301 ValAspPheGluHisLysAlaSerArgPheSerGlnGluHisGluGluMetAsnAlaLys 320
QY 1032 TTGGCGAATCAAGAAATCTCAACCGGCAACTTCGACTCAAACTGGTTGGCTTATCGCAA 1091
Db 321 LeuAlaAsnGlnGluSerHisAsnArgGlnLeuArgLeuLysLeuValGlyLeuThrGln 340
QY 1092 AGGATTGAGGAGCTGGAGAGACCAATAAAAAGCCTTCAGAAGGCAGAGGAAGCTCCAG 1151
Db 341 ArgIleGluGluLeuGluThrAsnLysAsnLeuGlnLysAlaGluGluLeuGln 360
QY 1152 GAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAACCTCCAGTCTCATGGCGGAAGTG 1211
Db 361 GluLeuArgAspLysIleAlaLysGlyGluCysGlyAsnSerSerLeuMetAlaGluVal 380
QY 1212 GAGAGTCTGCGCAAGCGGTGCTTGAGATGGAGGGCAAGGATGAAGAGATCACGAAGACC 1271
Db 381 GluAsnLeuArgLysArgValLeuGluMetGluGlyLysAspGluGluIleThrLysThr 400

QY 1272 GAGGCCAGTGCCGGGAGCTGAAGAAGAAGCTCCAAGAGGAAGAACACACAGCAAGGAA 1331
Db 401 GluSerGlnCysArgGluLeuArgLysLysLeuGlnGluGluHisSerLysGlu 420
QY 1332 CTTAGACTAGAAAGTGGAGAAAGCTGCAGAAGAGGATGTCTGAGCTGGAGAAGCTGGAGGAA 1391
Db 421 LeuArgLeuGluValGluLysLeuGlnLysArgMetSerGluLeuGluLysLeuGluGlu 440
QY 1392 GCGTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTCGAGAAGGAGAAG 1451
Db 441 AlaPheSerLysSerLysSerGluCysThrGlnLeuHisLeuAsnLeuGluLysGluLys 460
QY 1452 AACCTAACCAAGACCTGCTGAACGAGCTGGAGGTGTCAAAGAGTCGAGTTAAAGAACTC 1511
Db 461 AsnLeuThrLysAspLeuLeuAsnGluLeuGluValValLysSerArgValLysGluLeu 480
QY 1512 GAATGCTCCGAGAGTAGACTGGAGAAAGCCGAGTTAAGCCTCAAAAGATGACCTTACAAAG 1571
Db 481 GluCysSerGluSerArgLeuGluLysAlaGluLeuSerLeuLysAspAspLeuThrLys 500
QY 1572 CTGAAGTCCTTCACCTGTGATCTGTGTGATGAGAGGAAAAAATATGATGGAGAAAAATAAG 1631
Db 501 LeuLysSerPheThrValMetLeuValAspGluArgLysAsnMetMetGluLysIleLys 520
QY 1632 CAAGAAAGAGAGGAAAGTGGATGGGTTGAATAAAAACTTTAAAGTGGAGGAAAAAGTC 1691
Db 521 GlnGluGluArgLysValAspGlyLeuAsnLysAsnPheLysValGluGlnGlyLysVal 540
QY 1692 ATGGATGTGACGGAAAAAGCTAATCGAGGAAAAGCAAGAGCTTTTAAACTCAAACTCTGAA 1751
Db 541 MetAspValThrGluLysLeuIleGluGluSerLysLysLeuLeuLysLeuLysSerGlu 560
QY 1752 ATGGAGGAAAAAGGATFACAGTCTGCACAAAGGAGAGGATGAGCTGATGGGTAACCTGAGG 1811
Db 561 MetGluGluLysValTyrAsnLeuThrArgGluArgAspGluLeuIleGlyLysLeuLys 580
QY 1812 AGCGAAGAAAGGTCCTGTGTAACCTGAGCTGCAGTGCAGTACTTACTAAAAGACGGCTT 1871
Db 581 SerGluGluGluLysSerSerGluLeuSerCysSerValAspLeuLeuLysLysArgLeu 600
QY 1872 GATGGCATAGAGGAGTAGAAAGGGAATAAAACCGAGGTAGGTCTGTCGAAGGGGTCTGAG 1931
Db 601 AspGlyIleGluGluValGluArgGluIleThrArgGlyArgSerArgLysGlySerGlu 620
QY 1932 TTCACCTGCCCGGAAGACAATAAGATCAGAGAACTAACCGCTTGAATCGAGAGACTGAAG 1991
Db 621 LeuThrCysProGluAspAsnLysIleLysGluLeuThrLeuGluIleGluArgLeuLys 640
QY 1992 AAACGGCTCCAGCAGTTGGAGGTGGTGGAGGGGACTTGATGAAGACCGAGGACGAATAT 2051
Db 641 LysArgLeuGlnGlnLeuGluValValGluGlyAspLeuMetLysThrGluAspGluTyr 660
QY 2052 GACCAAGTTGGAGCAAGAGTTCAGAACCGAGCAGGATTAAGGCAAACTTCCTCTCCCAGCAG 2111
Db 661 AspGlnLeuGluGlnLysPheArgThrGluGlnAspLysAlaAsnPheLeuSerGlnGln 680
QY 2112 CTCGAGGAAATCAACACACCAAAATGGCCAAAGCACAAAGCCCATAGAGAAAGGGAGGCCGTG 2171
Db 681 LeuGluGluIleLysHisGlnIleAlaLysAsnLysAlaIleGluLysGlyGluValVal 700
QY 2172 AGCCAGGAAGCCGAACTGCGACACAGGTTTCGGCTGGAGGAGGCTAAAAGTCGTGATTTA 2231
Db 701 SerGlnGluAlaGluLeuArgHisArgPheArgLeuGluGluAlaLysSerArgAspLeu 720
QY 2232 CAGGCCGAGGTGCAGCTCTCAAGGAGAAGATCCACGAGCTGATGAACAAGGAAGACCAG 2291
Db 721 LysAlaGluValGlnAlaLeuLysGluLysIleHisGluLeuMetAsnLysGluAspGln 740
QY 2292 CTGCTCAGCTCCAAGTCGACTATTCGGTCCCTTCAGCAAAAGATTTATGGAAGGAAGAACT 2351
Db 741 LeuSerGlnLeuGlnValAspTyrSerValLeuGlnGlnArgPheMetGluGluAsn 760

QY	2352	AAGAACAAGAACATGGGGAGGAGGTCTCTCAATCTGACCAAGGAGCTAGAGCTTTCCAAG	2411
Db	761	LysAsnLysAsnMetGlyGlnGluValLeuAsnLeuThrLysGluLeuGluLeuSerLys	780
QY	2412	CGCTACAGCCGAGCTCTCAGGCCGAGTGGAAACGGCCGAAGGATGGTGGACGTGCTGTG	2471
Db	781	ArgTyrSerArgAlaLeuArgProSerValAsnGlyArgArgMetValAspValProVal	800
QY	2472	GCCTCCACTGGGGTGCAGACCGAGCGGTGTGCGGGGATGCTCGGAGGAGGAGACCCCG	2531
Db	801	ThrSerThrGlyValGlnThrAspAlaValSerGlyGluAlaAAGluGluThrPro	820
QY	2532	GCTGTGTTCAATTCGCAATCCTTCCAGGAGGAAAAATCACATCATGAGTAATCTTCACAG	2591
Db	821	AlaValPheIleArgLysSerPheGlnGluAsnHisIleMetSerAsnLeuArgGln	840
QY	2592	GTAGGCCTGAAGAAACCCATGAACCGGTCTCTCGGTCTCGACAGGTATCCCCCAGCAGCG	2651
Db	841	ValGlyLeuLysLysProValGluArgSerSerValLeuAspArgTyrProProAlaAla	860
QY	2652	AATGAGCTCACCATGAGGAAGTCTTGGATTCTTGGATGAGAAAAAGAGAAACCGTCTCT	2711
Db	861	AsnGluLeuThrMetArgLysSerTrpIleProTrpMetArgLysArgGluAsnGlyPro	880
QY	2712	TCCACTCCGAGGAGAAAGGCCCCAGGCCAAACACGAGGTGCAGGGCACCCCGGGAGCTG	2771
Db	881	SerIleThrGlnGluLysGlyProArgThrAsnSerSerProGlyHisProGlyGluVal	900
QY	2772	GTCCTAGCACCAAGCAGGGCCAGCCCTACACATCCGTGTGACACCAGATCATGAGAAC	2831
Db	901	ValLeuSerProLysGlnGlyGlnProLeuHisIleArgValThrProAspHisGluAsn	920
QY	2832	AGCACTGCCACCCCTGGAGATCACAAAGCCCCACATCTGAAGAGTTTTTCTCTAGTACCAC	2891
Db	921	SerThrAlaThrLeuGluIleThrSerProThrSerProThrSerGluGluPheSerSerThr	940
QY	2892	GTCATTCTCTACCTTAGCAACACAGAAACCAAGATAACCATATTATCCATCACCCCAATGTC	2951
Db	941	ValIleProThrLeuGlyAsnGlnLysProArgIleThrIleIleProSerProAsnVal	960
QY	2952	ATGTCGCAAAAGCCCCAAAGTGCAGATCCTACTCTCGGCCCAGAAACGAGCCATGTCCCT	3011
Db	961	MetProGlnLysGlnLysSerGlyAspThrThrLeuGlyProGluArgAlaMetSerPro	980
QY	3012	GTCACGATTACTACTATTTCCAGAGAGAAGAGCCCGGAAGGTGGAAGCGCCTTTGCC	3071
Db	981	ValThrIleThrThrPheSerArgGluLysThrProGluSerGlyArgGlyAlaPheAla	1000
QY	3072	GACAGGCCTGCATCCCCCATCCAAATCATGACGGTGTCACACATCTGCAGCTCCCACTGAA	3131
Db	1001	AspArgProThrSerProIleGlnIleMetThrValSerThrSerAlaAlaProAlaGlu	1020
QY	3132	ATCGCTGTCTCTCCTGAATCTCAGGAAGTGCCTATGGAAGGACTATCCTCAAAGTCACC	3191
Db	1021	IleAlaValSerProGluSerGlnGluMetProMetGlyArgThrIleLeuLysValThr	1040
QY	3192	CCGGAACAAACAACTGTTCCAGCCCCCGTGGGAAGTACAACTCCAATGTCTAATATCATC	3251
Db	1041	ProGluLysGlnThrValProThrProValArgLysTyrAsnSerAsnAlaAsnIleIle	1060
QY	3252	ACCACGGAAGACAATAAATTACATTACCTGGGTCTCAGTTTAAAGCGATCTCCTGGG	3311
Db	1061	ThrThrGluAspAsnLysIleHisIleHisLeuGlySerGlnPheLysArgSerProGly	1080
QY	3312	CCTGCCGTGAAGGCGTGAGCCCAAGTTATCACCCGTCCGGCCTGTCAACGTGACAGCGGAG	3371
Db	1081	ThrSerGlyGluGlyValSerProValIleThrValArgProValAsnValThrAlaGlu	1100
QY	3372	AAGGAGGTTTCTACAGGCACAGTCCCTTCGCTCTCCAGGAACCACTCTCTTCAAGACCC	3431
Db	1101	LysGluValSerThrGlyThrValLeuArgSerProArgAsnHisLeuSerSerArgPro	1120
QY	3432	GGTGCTAGCAAAAGTGACCAGCACTATAACTATAACCCCGGTCAACAACGTCTATCCACAG	3491

Db	1121	GlyAlaSerLysValThrSerThrIleThrIleThrProValThrThrSerSerAlaArg	1140
QY	3492	GGAACCCCAATCAGTGTCCAGGACAAGATGGGTCTATCTCAGCGGCTACCCCCACCCGATT	3551
Db	1141	GlyThrGlnSerValSerGlyGlnAspGlySerSerGlnArgProThrProThrArgIle	1160
QY	3552	CCTATGTCAAAAGGTATGAAAGCTGGAAGCCAGTAGTGGCAGCCTCAGGAGCAGGAAAT	3611
Db	1161	PrometSerLysGlyMetLysAlaGlyLysProValValAlaAlaProGlyAlaGlyAsn	1180
QY	3612	CTGACCAAAATTCAGCCTCGAGCTGAGACTCAGTCTATGAAAAATAGAGCTGAAGAAATCT	3671
Db	1181	LeuThrLysPheGluProArgAlaGluThrGlnSerMetLysIleGluLeuLysLysSer	1200
QY	3672	GCAGCCAGCAGCACTGCCTCTCTTTGGAGGGGGAAGGGC	3710
Db	1201	AlaAlaSerSerThrThrSerLeuGlyGlyGlyLysGly	1213
RESULT 3			
Q86TC3			
ID	Q86TC3	PRELIMINARY;	PRT; 1177 AA.
AC	Q86TC3;		
DT	01-JUN-2003	(TrEMBLrel. 24, Created)	
DT	01-JUN-2003	(TrEMBLrel. 24, Last sequence update)	
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)	
DE	Hypothetical protein DKF2p451B134.		
GN	DKF2P451B134.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,		
RA	Osanger A., Wiemann S.;		
RL	Submitted (APR-2003) to the EMBL/GenBank/DBDJ databases.		
DR	EMBL; AL832009; CAD89912.1; -		
DR	InterPro; IPR000533; Tropomyosin.		
DR	PRINTS; PR00194; TROPOMYOSIN.		
KW	Hypothetical protein.		
SQ	SEQUENCE	1177 AA; 134826 MW; D791D2A35FD10E4B	CRC64;
Alignment Scores:			
Pred. No.:	1.94e-261	Length:	1177
Score:	5456.50	Matches:	1086
Percent Similarity:	96.39%	Conservative:	36
Best Local Similarity:	93.30%	Mismatches:	41
Query Match:	70.58%	Indels:	1
DB:	4	Gaps:	1
US-10-788-793-1 (1-4364) x Q86TC3 (1-1177)			
QY	75	ATGAGATCACGAAATCAAGGTGGAGAAAGTTTCATCTAACGGGCATGTCTCTGCCCAAG	134
Db	1	MetArgSerArgAsnGlnGlyGlyGluSerAlaSerAspGlyHisIleSerCysProLys	20
QY	135	TCCTCCATCATCAGCAGTGGTGGTAAGGSCCCTCAGAAGATGCA---AAAAAGAAC	191
Db	21	ProSerIleIleGlyAsnAlaGlyGluLysSerLeuSerGluAspAlaLysLysArg	40
QY	192	AAGCCCAATCGGAAGGAGGAGGATGTCTATGGCTTCGGAACTATCAAAAGGCACCTCAA	251
Db	41	LysSerAsnArgLysGluAspAspValMetAlaSerGlyThrValLysArgHisLeuLys	60
QY	252	CCATCTGGAGAAAGTGAGAAAAAGACTAAGAAGTCTGTGGAGTTATCCAAGGAGACCTC	311
Db	61	ThrSerGlyGluCysGluArgLysThrLysLysSerLeuGluLeuSerLysGluAspLeu	80
QY	312	ATCCAGCTCCTGAGTATCATGGAAGGGGAGTTGTCAGGCTCGAGAAGATGTCTCCACATG	371
Db	81	IleGlnLeuLeuSerIleMetGluGlyGluLeuGlnAlaArgGluAspValIleHisMet	100

Db 821 AlaValPheIleArgLysSerPheGlnGluGluAsnHisIleMetSerAsnLeuArgGln 840

QY 2592 GTAGGCCTGAAGAAACCATGGAACGGTCTCGTCTCGACAGGTATCCCCAGCAGCG 2651

Db 841 ValGlyLeuLysLysProValGluArgSerSerValIleuAspArgTyrProProAlaAla 860

QY 2652 AATGAGCTCACCATGAGGAAGTCTTTGGATTCTTGGATGAGAAAAAGAGAAACGGTCT 2711

Db 861 AsnGluLeuThrMetArgLysSerTrpIleProTrpMetArgLysArgGluAsnGlyPro 880

QY 2712 TCCACTCCGCAGGAGAAAGGCCCCAGGCCAAACCCAGGGTGCAGGGCACCCCGGGAGCTG 2771

Db 881 SerIleThrGlnGluLysGlyProArgThrAsnSerSerProGlyHisProGlyGluVal 900

QY 2772 GTCCTAGCACCAAGCAGGCCCGCCCTACACATCCGTGTGACACCAGATCATGAGAAC 2831

Db 901 ValLeuSerProLysGlnGlyGlnProLeuHisIleArgValThrProAspHisGluAsn 920

QY 2832 AGCACTGCCACCCCTGGAGATCACAAAGCCCCACATCTGAAGAGTTTTTCTCTAGTACCAC 2891

Db 921 SerThrAlaThrLeuGluIleThrSerProThrSerGluGluPhePheSerThrThr 940

QY 2892 GTCATTCTTACCTTAGCAACACAGAAACCAAGATAACCATATTATCCATCACCAATGTC 2951

Db 941 ValIleProThrLeuGlyAsnGlnLysProArgIleThrIleIleProSerProAsnVal 960

QY 2952 ATGTCGCAAAAGCCCAAAAGTGCAGATCCTACTCTCGGCCCAGAACGAGCCATGTCCCCT 3011

Db 961 MetProGlnLysGlnLysSerGlyAspThrThrLeuGlyProGluArgAlaMetSerPro 980

QY 3012 GTCACGATTACTACTATTTCAGAGAGAAGAGCCCGGAAGGTGGAAGGCGCCTTTGCC 3071

Db 981 ValThrIleThrThrPheSerArgGluLysThrProGluSerGlyArgGlyAlaPheAla 1000

QY 3072 GACAGGCCTGCATCCCCCATCCAAATCATGACGGTGTCAACATCTGCAGCTCCCACTGAA 3131

Db 1001 AspArgProThrSerProIleGlnIleMetThrValSerThrSerAlaAlaProAlaGlu 1020

QY 3132 ATCGCTGTCTCTCTGAATCTCAGGAAGTGCCTATGGGAAGGACTATCTCAAAGTCACC 3191

Db 1021 IleAlaValSerProGluSerGlnGluMetProMetGlyArgThrIleLeuLysValThr 1040

QY 3192 CCGGAAAAACAACCTGTTCCAGCCCCCGTGGGAAGTACAACTCCAATGCTAATATCATC 3251

Db 1041 ProGluLysGlnThrValProThrProValArgLysTyrAsnSerAsnAlaAsnIleIle 1060

QY 3252 ACCACGGAAGACAATAAATTCACATTCCACTGGGTCTCAGTTTAAGCGATCTCCTGGG 3311

Db 1061 ThrThrGluAspAsnLysIleHisIleHisLeuGlySerGlnPheLysArgSerProGly 1080

QY 3312 CCTGCCGTGAAGCGGTGAGCCCGAGTTATCACCGTCCGGCCCTGTCAACGTGACAGCGGAG 3371

Db 1081 ThrSerGlyGluGlyValSerProValIleThrValArgProValAsnValThrAlaGlu 1100

QY 3372 AAGGAGGTTTCTACAGGCACAGTCTTCCGCTCTCCAGGAACCACTCTCTTCAAGACCC 3431

Db 1101 LysGlyValSerThrGlyThrValLeuArgSerProArgAsnHisLeuSerSerArgPro 1120

QY 3432 GGTGCTAGCAAAAGTGACAGCAGTCTTCCGCTCTCCAGGAACCACTCTCTTCAAGACCC 3491

Db 1121 GlyAlaSerLysValThrSerThrIleThrIleThrProValThrThrSerSerAlaArg 1140

QY 3492 GGAACCCCAATCAGTGTGAGGACAAGATGGGTCTATCTCAGCGGCCTACCCCGCCGATT 3551

Db 1141 GlyThrGlnSerValSerGlyGlnAspGlySerSerGlnArgProThrProThrArgIle 1160

QY 3552 CCTATGTCAAAA 3563

Db 1161 ProMetSerLys 1164

RESULT 4
Q9ULE5
ID Q9ULE5

PRELIMINARY; PRT; 1140 AA.

AC Q9ULE5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein KIAA1275 (Fragment).
GN KIAA1275.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
DR EMBL; AB033101; BAA86589.1; --
DR Genew; HGNC:21015; FILIP1.
DR InterPro; IPR000533; Tropomyosin.
DR PRINTS; PR00194; TROPOMYOSIN.
KW Hypothetical protein.
FT NON TER 1 1
SQ SEQUENCE 1140 AA; 130283 MW; 79EC2AE496556415 CRC64;

Alignment Scores:
Pred. No.: 3.9e-256 Length: 1140
Score: 5349.00 Matches: 1064
Percent Similarity: 96.65% Conservative: 32
Best Local Similarity: 93.83% Mismatches: 38
Query Match: 69.19% Indels: 0
DB: 4 Gaps: 0

US-10-788-793-1 (1-4364) x Q9ULE5 (1-1140)

QY 309 CTCATCCAGCTCCTGAGTATCATGGAAGGGGAGTTGCAGGCTCGAGAAGATGTCATCCAC 368

Db 7 IleValSerArgLeuIleLeuLeuHisPheLeuLeuGlnAlaArgGluAspValIleHis 26

QY 369 ATGCTGAGGACAGAGAAACCAAGCCCGAGGTCTTGGAGGCACACTATGGATCTGCAGAA 428

Db 27 MetLeuLysThrGluLysThrLysProGluValLeuGluAlaHisTyrGlySerAlaGlu 46

QY 429 CCTGAGAAAGTGCTTCGGGTCTTGCCACCGAGATGCCATCCTTGCTCAAGAGAAGTCCATA 488

Db 47 ProGluLysValIleuArgValLeuHisArgAspAlaIleLeuAlaGlnGluLysSerIle 66

QY 489 GGAGAAGACGCTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAAAAAGCAGAAG 548

Db 67 GlyGluAspValTyrGluLysProIleSerGluLeuAspArgLeuGluLysGlnLys 86

QY 549 GAGACGTACCGCCGATGCTAGAGCAGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 608

Db 87 GluThrTyrArgArgMetLeuGluGlnLeuLeuAlaGluLysCysHisArgArgThr 106

QY 609 GTGTACGAGCTGGAGAACGAGAACGACACACTGACTACTATGAAACAGACGACGACGAC 668

Db 107 ValTyrGluLeuGluAsnGluLysHisLysHisThrAspTyrMetAsnLysSerAspAsp 126

QY 669 TTCACCAACCTGCTGGAGCAGGAGCGAGAGGTTGAAAAAGCTCCTTGAACAGAAAAA 728

Db 127 PheThrAsnLeuLeuGluGlnGluArgGluArgLeuLysLysLeuLeuGlnGluLys 146

QY 729 GCTTACCAAGCCGCAAGAAAAAGGAAACGCTAAGCGGCTCAACAACTTCGAGATGAG 788

Db 147 AlaTyrGlnAlaArgLysGluLysGluAsnAlaLysArgLeuAsnLysLeuArgAspGlu 166

QY 789 CTTGTGAAGCTCAAGTCTTCGCCCTCATGTTGGTGGACGAGGCGCAGATGCACATCGAG 848

Db 167 LeuValLysLeuLysSerPheAlaLeuMetLeuValAspGluArgGlnMetHisIleGlu 186

QY	849	CAACTGGGCGCTGCAGAGTCAGAAAAGTCCAGGACCTCACTCAGAAAGCTGAGGGAGGAGGAA	908
Db	187	GlnLeuGlyLeuGlnSerGlnLysValGlnAspLeuThrGlnLysLeuArgGluGlu	206
QY	909	GA AAAA CTCAAAGCGGTCACTTACAAATCCAAAGGAAGACCGCCAGAAAGCTGCTCAAGTTA	968
Db	207	GluLysLeuLysAlaIleThrSerLysSerLysGluAspArgGlnLysLeuLysLeu	226
QY	969	GAAGTGGACTTCGAACACAAAGGCCTCGAGGTTTTCCAGGAGCACGAAGAGATGAACGCC	1028
Db	227	GluValaspPheGluHisLysAlaSerArgPheSerGlnGluHisGluGluMetAsnAla	246
QY	1029	AAATTGGCGAATCAAGAATCTCACAACCGCGCAACTTCGACTCAAACCTGGTTGGCTTATCG	1088
Db	247	LysLeuAlaAsnGlnGluSerHisAsnArgGlnLeuArgLeuLysLeuValGlyLeuThr	266
QY	1089	CAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAAGGCAGAGGAGAGCTC	1148
Db	267	GlnArgIleGluGluLeuGluGluThrAsnLysAsnLysAlaGluGluLeu	286
QY	1149	CAGGAGCTGAGAGAGAAAATTGCCAAAGGGGAATGTGGAACCTCCAGTCTCATGGCGAA	1208
Db	287	GlnGluLeuArgAspLysIleAlaLysGlyGluCysGlyAsnSerSerLeuMetAlaGlu	306
QY	1209	GTGAGAGTCTGCGCAAGCGCGTGTCTGAGATGGAGGCAAGGATGAAGAGATCACGAAG	1268
Db	307	ValGluAsnLeuArgLysArgValLeuGluMetGluGlyLysAspGluGluIleThrLys	326
QY	1269	ACCGAGGCCAGTGCCGGGAGCTGAAGAAGAAGCTCCAAGAGGAAGAACACCACAGCAAG	1328
Db	327	ThrGluSerGlnCysArgGluLeuArgLysLysLeuGlnGluGluGluHisSerLys	346
QY	1329	GAACTTAGACTAGAAAGTGGAGAAGCTGCAGAAGAGGATGTCTGAGCTGGAGAAGCTGGAG	1388
Db	347	GluLeuArgLeuGluValGluLysLeuGlnLysArgMetSerGluLeuLysLeuGlu	366
QY	1389	GAAGCGTTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAGGAG	1448
Db	367	GluAlaPheSerLysSerLysSerGluCysThrGlnLeuHisLeuAsnLeuGluLysGlu	386
QY	1449	AAGAACTTAACCAAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAA	1508
Db	387	LysAsnLeuThrLysAspLeuLeuAsnGluLeuGluValValLysSerArgValLysGlu	406
QY	1509	CTCGAATGCTCCGAGAGTAGACTGGAGAAGCCCGAGTTAAGCCTCAAGATGACCTTACA	1568
Db	407	LeuGluCysSerGluSerArgLeuGluLysAlaGluLeuSerLeuLysAspLeuThr	426
QY	1569	AAGCTAAGTCTTCACTGTGTGATGCTGGTGGATGAGAGGAAAAATATGATGGAGAAAATA	1628
Db	427	LysLeuLysSerPheThrValMetLeuValAspGluArgLysAsnMetMetGluLysIle	446
QY	1629	AAGCAAGAAGAGAGGAAAGTGGATGGGTTGAATAAAAACTTTAAGGTGGAGCAGGGAAAA	1688
Db	447	LysGlnGluGluArgLysValAspGlyLeuAsnLysAsnPheLysValGluGlnGlyLys	466
QY	1689	GTCATGGATGTGACGGAAAAAGCTAATCGAGGAAAGCAAGAAGCTTTTAAACTCAATCT	1748
Db	467	ValMetAspValThrGluLysLeuIleGluGluSerLysLysLeuLeuLysLysSer	486
QY	1749	GAATGGAGGAAAAAGGAGTACAGTCTGACAAAGGAGAGGGATGAGCTGATGGGTAAACTG	1808
Db	487	GluMetGluGluLysValTyrAsnLeuThrArgGluArgAspGluLeuIleGlyLysLeu	506
QY	1809	AGGAGCGAAGAAGAGGTCTGTGAACTGAGCTCGAGTGTAGACTTACTAAAGAGCGG	1868
Db	507	LysSerGluGluGluLysSerSerGluLeuSerCysSerValAspLeuLeuLysLysArg	526
QY	1869	CTTGATGGCATAGAGGAGGTAGAAAGGAAAAATAAACCGAGGTAGTTCGTCAAGGGTCT	1928
Db	527	LeuAspGlyIleGluGluValGluArgGluIleThrArgGlyArgSerArgLysGlySer	546
QY	1929	GAGTTACCTGCCCGGAAGACAATAAGATCAGAGAACTAAACGCTTGAANAATCGAGAGACTG	1988

Db	547	GluLeuThrCysProGluAspAsnLysIleLysGluLeuThrLeuGluIleGluArgLeu	566
QY	1989	AAGAAACGGCTCCAGCAGTTGGAGTGGTGGAGGGGACTTGTATGAAGACCGAGGACGAA	2048
Db	567	LysLysArgLeuGlnGlnLeuGluValValGluGlyAspLeuMetLysThrGluAspGlu	586
QY	2049	TATGACCAGTTGGAGCAGAAAGTTCAGAACCGAGCAGGATAAAGGCAAACTTCTCTCCCAG	2108
Db	587	TyrAspGlnLeuGluGlnLysPheArgThrGluGlnAspLysAlaAsnPheLeuSerGln	606
QY	2109	CAGCTCGAGGAAATCAAAACACCAATGGCCAAAGCACAAGCCATAGAGAAAAGGGAGGCC	2168
Db	607	GlnLeuGluGluIleLysHisGlnIleAlaLysAsnLysAlaIleGluLysGlyGluVal	626
QY	2169	GTGAGCCAGGAAGCCGAACTGCGACACAGGTTTCGGCTGGAGGAGGCTAAAAAGTCGTGAT	2228
Db	627	ValSerGlnGluAlaGluLeuArgHisArgPheArgLeuGluGluAlaLysSerArgAsp	646
QY	2229	TTACAGGCCGAGGTGCAGGCTCTCAAGGAGAAAGATCCACCGAGCTGATGAACAAGGAAGAC	2288
Db	647	LeuLysAlaGluValGlnAlaLeuLysGluLysIleHisGluLeuMetAsnLysGluAsp	666
QY	2289	CAGCTGTCTCAGCTCCAAGTCGACTATTCCGTCCTTCAGCAAAAGATTTATGGAAGAAGAA	2348
Db	667	GlnLeuSerGlnLeuGlnValAspTyrSerValLeuGlnGlnArgPheMetGluGluGlu	686
QY	2349	ACTAAGAAACAAGACATGGGGAGGAGGTCTCTCAATCTGACCAAGGAGCTAGAGCTTTCC	2408
Db	687	AsnLysAsnLysAsnMetGlyGlnGluValLeuAsnLeuThrLysGluLeuGluLeuSer	706
QY	2409	AAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGGAAGGATGGTGGACGTGCCT	2468
Db	707	LysArgTyrSerArgAlaLeuArgProSerValAsnGlyArgArgMetValAspValPro	726
QY	2469	GTGGCTCCACTGGGGTGCAGACCAGAGCGGTGTGCGGGGATGCTCGGGAGGAGGAGACC	2528
Db	727	ValThrSerThrGlyValGlnThrAspAlaValSerGlyGluAlaAlaGluGluGluThr	746
QY	2529	CCGGCTGTGTTTCATTTCGAAAATCCTTCAGGAGGAAAAATCACATCATGAGTAATCTTCGA	2588
Db	747	ProAlaValPheIleArgLysSerPheGlnGluGlnAsnHisIleMetSerAsnLeuArg	766
QY	2589	CAGTAGGCCTGAAGAAACCCATCGAACCGGTCTCTCGTCTCTCGACAGGTATCCCCAGCA	2648
Db	767	GlnValGlyLeuLysLysProValGluArgSerSerValLeuAspArgTyrProProAla	786
QY	2649	GCGAATGAGCTCACCATGAGGAAGTCTTGGATTCTCTGGATGAGAAAAAGAGAAAAACGGT	2708
Db	787	AlaAsnGluLeuThrMetArgLysSerTrpIleProTrpMetArgLysArgGluAsnGly	806
QY	2709	CTTCCACTCCGCAGGAGAAAGGGCCAGGCCAAACACGAGGTGACGGGCACCCGGGGAG	2768
Db	807	ProSerIleThrGlnGluLysGlyProArgThrAsnSerSerProGlyHisProGlyGlu	826
QY	2769	CTGCTCCTAGCACCAAGCAGGGCCAGCCCTACACATCCGTGTGACACCAGATCATGAG	2828
Db	827	ValValLeuSerProLysGlnGlnProLeuHisIleArgValThrProAspHisGlu	846
QY	2829	AAACAGCACTGCCACCTGGAGATCAAAAGCCCAATCTGAAGAGTTTCTCTAGTACC	2888
Db	847	AsnSerThrAlaThrLeuGluIleThrSerProThrSerGluGluPhePheSerSerThr	866
QY	2889	ACCGTCATTCTTACCTTAGGCAACCAAGAAACCAAGAAATAACCATTTATTCATCACC	2948
Db	867	ThrValIleProThrLeuGlyAsnGlnLysProArgIleThrIleProSerProAsn	886
QY	2949	GTGATGTCGAAAAGCCCAAAAGTCAGATCTTACTCTCGGCCAGAAACGAGCCATGTCC	3008
Db	887	ValMetProGlnLysGlnLysSerGlyAspThrThrLeuGlyProGluArgAlaMetSer	906
QY	3009	CCGTGCACGATTACTACTATTTCAGAGAGAGAGCCCGGAAGGTGGGAAGGAGCGCCTTT	3068

Db	907	ProValThrIleThrThrPheSerArgGluLysThrProGluSerGlyArgGlyAlaPhe	926
Qy	3069	GCCGACAGGCCTGCATCCCCCATCCAAATCATGACGGTGTCAACATCTGCAGCTCCCACT	3128
Db	927	AlaAspArgProThrSerProIleGlnIleMetThrValSerThrSerAlaAlaProAla	946
Qy	3129	GAATCGCTGTCTCTCTGAATCTCAGGAAGTGCCTATGGGAAGGACTATCCTCAAAGTC	3188
Db	947	GluIleAlaValSerProGluSerGlnGluMetProMetGlyArgThrIleLeuLysVal	966
Qy	3189	ACCCCGGAAAAACAAACTGTTCCAGCCCCCGTGCGGGAAGTACAACCTCCAATGCTAATATC	3248
Db	967	ThrProGluLysGlnThrValProThrProValArgLysTyrAsnSerAsnAlaAsnIle	986
Qy	3249	ATCACCACGGAAGACAAATAAAATTCACATTACCTGGGTTCTCAGTTTAAGCGATCTCCT	3308
Db	987	IleThrThrGluAspAsnLysIleHisIleHisLeuGlySerGlnPheLysArgSerPro	1006
Qy	3309	GGGCCTGCCGCTGAAGGCGTGAGCCCGATTATCACCGTCCGGCCTGTCAACGTGACAGCG	3368
Db	1007	GlyThrSerGlyGluGlyValSerProValIleThrValArgProValAsnValThrAla	1026
Qy	3369	GAGAAGGAGGTTTCTACAGGCACAGTCTTCGCTCTCCAGGAACCACTCTCTTCAAGA	3428
Db	1027	GluLysGluValSerThrGlyThrValLeuArgSerProArgAsnHisLeuSerSerArg	1046
Qy	3429	CCCGTGTCTAGCAAAAGTGACCAGCACTATAACTATAACCCCGGTCAACCGTCATCCACA	3488
Db	1047	ProGlyAlaSerLysValThrSerThrIleThrIleThrProValThrThrSerSerAla	1066
Qy	3489	CGAGGAACCCAAATCAGTGTGAGGACAAGATGGGTCTATCTCAGCGGCCTACCCCAACCCGC	3548
Db	1067	ArgGlyThrGlnSerValSerGlyGlnAspGlySerSerGlnArgProThrProThrArg	1086
Qy	3549	ATTCTATGTCAAAAGGTATGAAAGCTGGAAGCCAGTAGTGGCAGCCTCAGGAGCAGGA	3608
Db	1087	IleProMetSerLysGlyMetLysAlaGlyLysProValValAlaAlaProGlyAlaGly	1106
Qy	3609	AATCTGACCAAATTCAGGCCTCGAGCTGAGACTCAGTCTATGAAAATAGAGCTGAAGAAA	3668
Db	1107	AsnLeuThrLysPheGluProArgAlaGluThrGlnSerMetLysIleGluLeuLysLys	1126
Qy	3669	TCTGCAGCCAGCAGCACTGCCTCTCTTGGAGGGGGAAGGGC	3710
Db	1127	SerAlaAlaSerSerThrThrSerLeuGlyGlyGlyLysGly	1140
RESULT 5			
Q8JZS5			
ID	Q8JZS5	PRELIMINARY;	PRT; 965 AA.
AC	Q8JZS5;		
DT	01-OCT-2002	(TrEMBLrel. 22, Created)	
DT	01-OCT-2002	(TrEMBLrel. 22, Last sequence update)	
DT	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)	
DE	Filamin-interacting protein S-FILIP.		
GN	S-FILIP.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Wistar; TISSUE=Forebrain;		
RX	MEDLINE=22100428; PubMed=12055638;		
RA	Nagano T., Yoneda T., Hatanaka Y., Kubota C., Murakami F., Sato M.;		
RT	"Filamin A-interacting protein (FILIP) regulates cortical cell		
RT	migration out of the ventricular zone.";		
RL	Nat. Cell Biol. 4:495-501(2002).		
DR	EMBL; D87257; BAC00852.1; -.		
DR	InterPro; IPR000533; Tropomyosin.		
DR	PRINTS; PR00194; TROPOMYOSIN.		
SQ	SEQUENCE 965 AA; 109188 MW; 50352DF4F8236B40 CRC64;		

Alignment Scores:

Pred. No.:	6.3e-231	Length:	965
Score:	4838.00	Matches:	965
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	62.58%	Indels:	0
DB:	11	Gaps:	0
US-10-788-793-1 (1-4364) x Q8JZS5 (1-965)			
QY	816	ATGTTGGTGGACGACGAGGCAGATGCACATCGAGCAACTGGGCCTGCAGAGTCAGAAAATC	875
Db	1	MetLeuValAspGluArgGlnMetHisIleGluGlnLeuGlyLeuGlnSerGlnLysVal	20
QY	876	CAGGACCTCACTCAGAAGCTGAGGGAGGAGGAGAGAAAACCTCAAAGCGGTCACTTACAAA	935
Db	21	GlnAspLeuThrGlnLysLeuArgGluGluGluLysLeuLysAlaValThrTyrLys	40
QY	936	TCCAAGGAAGACCCGACAGAGCTGCTCAAGTTAGAGTGGACTTCGAACACAAAGGCCTCG	995
Db	41	SerLysGluAspArgGlnLysLeuLeuLysLeuGluValAspPheGluHisLysAlaSer	60
QY	996	AGGTTTCCCAGGACGACGAAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCACAAAC	1055
Db	61	ArgPheSerGlnGlnHisGluGluMetAsnAlaLysLeuAlaAsnGlnGluSerHisAsn	80
QY	1056	CGGCAACTTCGACTCAAACCTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGAGAGACC	1115
Db	81	ArgGlnLeuArgLeuLysLeuValGlyLeuSerGlnArgIleGluGluLeuGluThr	100
QY	1116	AATAAAGCCTTCAGAAAGCAGAGGAAAGAGTCCAGGAGCTGAGAGAGAAAATTGCCAAA	1175
Db	101	AsnLysSerLeuGlnLysAlaGluGluGluGlnGluLeuArgGluLysIleAlaLys	120
QY	1176	GGGGAATGTGAAACTCCAGTCTCATGTCGCGAAGTGGAGAGTCTGCGCAAGCGCGTCTT	1235
Db	121	GlyGluCysGlyAsnSerSerLeuMetAlaGluValGluSerLeuArgLysArgValLeu	140
QY	1236	GAGTGGAGGGCAAGGATGAAGAGATCACGAAGACCCGAGGCCAGTCCGGGAGCTGAAG	1295
Db	141	GluMetGluGlyLysAspGluGluIleThrLysThrGluAlaGlnCysArgGluLeuLys	160
QY	1296	AAGAAGCTCCAAGAGGAACACACCACAGCAAGGAACTTAGACTAGAAAGTGGAGAAGCTG	1355
Db	161	LysLysLeuGlnGluGluHisHisSerLysGluLeuArgLeuGluValGluLysLeu	180
QY	1356	CAGAAGAGGATGTCTGAGCTGGAGAAGCTGGAGGAAGCGTTACGCCGGAGTAAGTCGGAA	1415
Db	181	GlnLysArgMetSerGluLeuGluLysLeuGluAlaPheSerArgSerLysSerGlu	200
QY	1416	TGCACCCAGTCCATCTGAACCTGGAGAGGAGAGAAACCTAACCAAAGACCTGCTGAAC	1475
Db	201	CysThrGlnLeuHisLeuAsnLeuGluLysGluLysAsnLeuThrLysAspLeuLeuAsn	220
QY	1476	GAGCTGGAGGTGCTCAAGAGTTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAG	1535
Db	221	GluLeuGluValValLysSerArgValLysGluLeuGluCysSerGluSerArgLeuGlu	240
QY	1536	AAGCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAGTCCCTTCACTGTGATGCTG	1595
Db	241	LysAlaGluLeuSerLeuLysAspAspLeuThrLysLeuLysSerPheThrValMetLeu	260
QY	1596	GTGGATGAGAGGAAAAATATGATGGAGAAAATAAAGCAAGAGAGAGGAAAGTGGATGGG	1655
Db	261	ValAspGluArgLysAsnMetMetGluLysIleLysGlnGluGluArgLysValAspGly	280
QY	1656	TTGAATAAAAACTTTAAGGTGGAGCAGGGAAAAAGTTCATGGATGTGACGGAAAAAGCTAATC	1715
Db	281	LeuAsnLysAsnPheLysValGluGlnGlyLysValMetAspValThrGluLysLeuIle	300
QY	1716	GAGGAAAGCAAGAAGCTTTTAAACCTCAAATCTGAAATCGAGGAAAAGGAGTACAGTCTG	1775
Db	301	GluGluSerLysLysLeuLeuLysLeuLysSerGluMetGluGluLysGluTyrSerLeu	320

QY	1776	ACAAAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGGAAGAAAGGTCCTGTGAA	1835
Db	321	ThrLysGluArgAspGluLeuMetGlyLysLeuArgSerGluGluArgSerCysGlu	340
QY	1836	CTGAGCTCAGTGTAGACTTACTAAAGAACGGCTTGATGGCATAGAGAGGTTAGAAAGG	1895
Db	341	LeuSerCysSerValAspLeuLeuLysLysArgLeuAspGlyIleGluGluValGluArg	360
QY	1896	GAATATAACCGAGGTAGTCGTGCAAGGGTCTTGAGTTACCTGCCCGAAGACAAATAAG	1955
Db	361	GluIleAsnArgGlyArgSerCysLysGlySerGluPheThrCysProGluAspAsnLys	380
QY	1956	ATCAGAGAACTAACGCTTGAAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTG	2015
Db	381	IleArgGluLeuThrLeuGluIleGluArgLeuLysLysArgLeuGlnGlnLeuGluVal	400
QY	2016	GTGGAGGGGACTTGATGAAGACCCGAGGACGAATATGACCAGTTGGAGCAGAAGTTCAGA	2075
Db	401	ValGluGlyAspLeuMetLysThrGluAspGluTyrAspGlnLeuGluGlnLysPheArg	420
QY	2076	ACCGAGCAGGATAAGGCAAACTTCCTCTCCAGCAGCTCGAGGAAATCAAAACACCAAATG	2135
Db	421	ThrGluGlnAspLysAlaAsnPheLeuSerGlnGlnLeuGluGluIleLysHisGlnMet	440
QY	2136	GCCAAAGCACAAAGCCATAGAGAAAGGGGAGGCCGTGAGCCAGGAAGCCGAACTGCCGACAC	2195
Db	441	AlaLysHisLysAlaIleGluLysGlyGluAlaValSerGlnGluAlaGluLeuArgHis	460
QY	2196	AGGTTTCGGCTGGAGGAGGCTAAAAGTCGTGATTTACAGGCCGAGGTCAGGCTCTCAAG	2255
Db	461	ArgPheArgLeuGluAlaLysSerArgAspLeuGlnAlaGluValGlnAlaLeuLys	480
QY	2256	GAGAAGATCCACGAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTAT	2315
Db	481	GluLysIleHisGluLeuMetAsnLysGluAspGlnLeuSerGlnLeuGlnValAspTyr	500
QY	2316	TCGGTCCTTCAGCAAAAGATTATGGAAGAAAGAACTAAGAAACAAGAACATGGGAGGAG	2375
Db	501	SerValLeuGlnGlnArgPheMetGluGluGluThrLysAsnLysAsnMetGlyArgGlu	520
QY	2376	GTCCTCAATCTGACCAAGGAGTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCG	2435
Db	521	ValLeuAsnLeuThrLysGluLeuGluLeuSerLysArgTyrSerArgAlaLeuArgPro	540
QY	2436	AGTGGGAACGGCCGAGGATGGTGGACGTGCCTGTGCGCTCCACTGGGGTGCAGACCGAG	2495
Db	541	SerGlyAsnGlyArgArgMetValAspValProValAlaSerThrGlyValGlnThrGlu	560
QY	2496	GCGGTGTGCGGGATGCTGCGGAGGAGAGACCCCGGTGTGTTCAFTTCGAAATCCTTC	2555
Db	561	AlaValCysGlyAspAlaAlaGluGluGluThrProAlaValPheIleArgLysSerPhe	580
QY	2556	CAGGAGGAAATCACATCATGAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCATGGAA	2615
Db	581	GlnGluGluAsnHisIleMetSerAsnLeuArgGlnValGlyLeuLysLysProMetGlu	600
QY	2616	CGGTCCCTCGGTCCCTCGACAGGTATCCCCCAGCAGCGGAATGAGCTCACCATGAGGAAGTCT	2675
Db	601	ArgSerSerValLeuAspArgTyrProProAlaAlaAsnGluLeuThrMetArgLysSer	620
QY	2676	TGGATTCCCTTGGATGAGAAAAAGAGAAAAACGGTCCCTTCCACTCCGCAAGGAGAAAGGCC	2735
Db	621	TrpIleProTrpMetArgLysArgGluAsnGlyProSerThrProGlnGluLysGlyPro	640
QY	2736	AGGCCAAACCAAGGTCAGGGCACCCCGGGAGCTGCTCCTAGCACCAAGCAGGGCCAG	2795
Db	641	ArgProAsnGlnGlyAlaGlyHisProGlyGluLeuValLeuAlaProLysGlnGlyGln	660
QY	2796	CCCTTACACATCCGTGTGACACCAGATCATGAGAACAGCACTGCCACCTGGAGATCACA	2855
Db	661	ProLeuHisIleArgValThrProAspHisGluAsnSerThrAlaThrLeuGluIleThr	680
QY	2856	AGCCCCACATCTGAAGAGTTTTTCTCTAGTACCACCGTCATTCCTACCTTAGGCAACCAG	2915

Db	681	SerProThrSerGluGluPhePheSerSerThrThrValIleProThrLeuGlyAsnGln	700
QY	2916	AAACCAAGAATAACCATTTATCCATCACCCTCATGTGTATGTTCGCAAAAGCCCAAAGTGCA	2975
Db	701	LysProArgIleThrIleIleProSerProAsnValMetSerGlnLysProLysSerAla	720
QY	2976	GATCCTACTCTCGGCCCAGAAACGAGCCATGTCCCCTGTGCACGATTACTATTTCACGA	3035
Db	721	AspProThrLeuGlyProGluArgAlaMetSerProValThrIleThrThrIleSerArg	740
QY	3036	GAGAGAGCCCGGAAGGTGGAAGGCGCCTTTTCCGACAGGCCTGCATCCCCCATCCAA	3095
Db	741	GluLysSerProGluGlyArgSerAlaPheAlaAspArgProAlaSerProIleGln	760
QY	3096	ATCATGACGGTGTCAACATCTGCAGTCCCACCTGAAATCGCTGTCTCTCCTGAATCTCAG	3155
Db	761	IleMetThrValSerThrSerAlaAlaProThrGluIleAlaValSerProGluSerGln	780
QY	3156	GAAGTGCCCTATGGGAAGGACTATCCTCAAAGTCAACCCCGGAAAAACAACACTGTTCCAGCC	3215
Db	781	GluValProMetGlyArgThrIleLeuLysValThrProGluLysGlnThrValProAla	800
QY	3216	CCCGTCGGGAAGTACAACCTCCAATGCTAATATCATCACACGGAAGACAATAAAATTACAC	3275
Db	801	ProValArgLysTyrAsnSerAsnAlaAsnIleIleThrThrGluAspAsnLysIleHis	820
QY	3276	ATTCACCTGGTTCTCAGTTTAAGCGATCTCCTGGGCCTGCCGTGAAGCGGTGAGCCCA	3335
Db	821	IleHisLeuGlySerGlnPheLysArgSerProGlyProAlaAlaGluGlyValSerPro	840
QY	3336	GTTATCACCGTCGCGCCTGTCAACGTGACAGCGGAGAAGGAGGTTTCTACAGGCACAGTC	3395
Db	841	ValIleThrValArgProValAsnValThrAlaGluLysGluValSerThrGlyThrVal	860
QY	3396	CTTCGCTCTCCCAGAACCACTCTCTTCAAGACCCGGTGTAGCAAAGTGACCAGCACT	3455
Db	861	LeuArgSerProArgAsnHisLeuSerSerArgProGlyAlaSerLysValThrSerThr	880
QY	3456	ATAACTATAACCCCGGTCACAACGTTCATCCACACAGGAACCCCAATCAGTGTTCAGGACAA	3515
Db	881	IleThrIleThrProValThrThrSerSerThrArgGlyThrGlnSerValSerGlyGln	900
QY	3516	GATGGGTCACTCAGCGGCCTACCCCAACCCGCAATTCCTATGTCAAAAGGTATGAAAGCT	3575
Db	901	AspGlySerSerGlnArgProThrProThrArgIleProMetSerLysGlyMetLysAla	920
QY	3576	GGAAAGCCAGTAGTGGCAGCCTCAGGAGCAGGAATCTGACCAATTCACGCTCGAGCT	3635
Db	921	GlyLysProValValAlaAlaSerGlyAlaGlyAsnLeuThrLysPheGlnProArgAla	940
QY	3636	GAGACTCAGTCTATGAAAAATAGAGCTGAAGAAATCTGCAGCCAGCAGCACTGCCTCTCTT	3695
Db	941	GluThrGlnSerMetLysIleGluLeuLysLysSerAlaAlaSerThrAlaSerLeu	960
QY	3696	GGAGGGGGGAAGGC 3710	
Db	961	GlyGlyGlyLysGly 965	
RESULT 6			
ID	Q96SK6	PRELIMINARY;	PRT; 965 AA.
AC	Q96SK6;		
DT	01-DEC-2001	(TrEMBLrel. 19, Created)	
DT	01-DEC-2001	(TrEMBLrel. 19, Last sequence update)	
DT	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)	
DE	Hypothetical protein FLJ14799.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		

Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,
RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,
RA Masuho Y., Kanehori K.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBDJ databases.
DR EMBL; AK027705; BAB55310.1; -;
DR InterPro; IPR000533; Tropomyosin.
DR PRINTS; PR00194; TROPOMYOSIN.
KW Hypothetical protein.
SQ SEQUENCE 965 AA; 109200 MW; 25EB750B47E42433 CRC64;

Alignment Scores:
Pred. No.: 2.84e-216 Length: 965
Score: 4541.00 Matches: 904
Percent Similarity: 96.58% Conservative: 28
Best Local Similarity: 93.68% Mismatches: 33
Query Match: 58.74% Indels: 0
DB: 4 Gaps: 0

US-10-788-793-1 (1-4364) x Q96SK6 (1-965)

QY 816 ATGTTGGTGACGAGGCGAGATGCACATCGAGCAACTGGGCCTGCAGAGTCAGAAAGTC 875
Db 1 MetLeuValAspGluArgGlnMetHisIleGluGlnLeuGlyLeuGlnSerGlnLysVal 20

QY 876 CAGGACCTCACTCAGAACTGAGGAGGAGGAAGAAAACTCAAAGCGGTCACTTACAAA 935
Db 21 GlnAspLeuThrGlnLysLeuArgGluGluGluLysLeuLysAlaIleThrSerLys 40

QY 936 TCCAAGGAAGACCGCAGAAAGCTGCTCAAGTTAGAGTGGAATTCGAACACAGGCCTCG 995
Db 41 SerLysGluAspArgGlnLysLeuLeuLysLeuGluValAspPheGluHisLysAlaSer 60

QY 996 AGGTTTTCCAGGAGCAGGAAGAGATGAACGCCAAATTGGCGAATCAAGAAATCTCACAAC 1055
Db 61 ArgPheSerGlnGluHisGluGluMetAsnAlaLysLeuAlaAsnGlnGluSerHisAsn 80

QY 1056 CGGCAACTTCGACTCAAACTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGAAGAGACC 1115
Db 81 ArgGlnLeuArgLeuLysLeuValGlyLeuThrGlnArgIleGluGluLeuGluThr 100

QY 1116 AATAAAGCCTTCAGAAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAATTGCCAAA 1175
Db 101 AsnLysAsnLeuGlnLysAlaGluGluGluLeuGlnGluLeuArgAspLysIleAlaLys 120

QY 1176 GGGGAATGTGGAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTGCTT 1235
Db 121 GlyGluCysGlyAsnSerSerLeuMetAlaGluValGluAsnLeuArgLysArgValLeu 140

QY 1236 GAGATGGAGGGCAAGATGAAGAGATCACGAAGACCGAGGCCAGTGCCGGGAGCTGAAG 1295
Db 141 GluMetGluGlyLysAspGluGluIleThrLysThrGluSerGlnCysArgGluLeuArg 160

QY 1296 AAGAAGCTCCAAGAGGAAGAACACCACAGCAAGGAACTTAGACTAGAAAGTGGAGAAAGCTG 1355
Db 161 LysLysLeuGlnGluGluGluHisHisSerLysLysGluLeuArgLeuGluValGluLysLeu 180

QY 1356 CAGAAGAGGATGTCTGAGCTGGAGAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAA 1415
Db 181 GlnLysArgMetSerGluLeuGluLysLeuGluAlaPheSerLysSerLysSerGlu 200

QY 1416 TGCACCCAGCTCCATCTGAACCTGGAGAAGGAGAAGAACCTTAACCAAGACCTGCTGAAC 1475
Db 201 CysThrGlnLeuHisLeuAsnLeuGluLysGluLysAsnLeuThrLysAspLeuLeuAsn 220

QY 1476 GAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAG 1535
Db 221 GluLeuGluValValLysSerArgValLysGluLeuGluCysSerGluSerArgLeuGlu 240

QY 1536 AAGGCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAGTCCCTTCACTGTGATGCTG 1595
Db 1 ArgSerSerValLeuAspArgTyrProProAlaAlaAsnGluLeuThrMetArgLysSer 620

Db 241 LysAlaGluLeuSerLeuLysAspAspLeuThrLysLeuLysSerPheThrValMetLeu 260

QY 1596 GTGGATGAGAGGAAAAAATATGATGGAGAAAAATAAAGCAAGAGAGAGGAAAGTGGATGGG 1655
Db 261 ValAspGluArgLysAsnMetMetGluLysIleLysGlnGluArgLysValAspGly 280

QY 1656 TTGAATAAAAACTTTAAGGTGGAGCAGGGGAAAAAGTCATGGATGTGACCGGAAAAAGCTAATC 1715
Db 281 LeuAsnLysAsnPheLysValGluGlnGlyLysValMetAspValThrGluLysLeuIle 300

QY 1716 GAGGAAAGCAAGAGCTTTTAAAACTCAAATCTGAAATGGAGGAAAAAGGAGTACAGTCTG 1775
Db 301 GluGluSerLysLysLeuLeuLysLeuLysSerGluMetGluGluLysValTyrAsnLeu 320

QY 1776 ACAAAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGAAGAAGAAAGTCTCTGTGAA 1835
Db 321 ThrArgGluArgAspGluLeuIleGlyLysLeuLysSerGluGluLysSerSerGlu 340

QY 1836 CTGAGCTGCAGTGTAGACTTACTAAAGAAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGG 1895
Db 341 LeuSerCysSerValAspLeuLeuLysLysArgLeuAspGlyIleGluGluValGluArg 360

QY 1896 GAAATAAACCGAGGTAGGTCGTGCAAGGGGTCTGAGTTCACCTGCCCGGAAGACAATAAG 1955
Db 361 GluIleThrArgGlyArgSerArgLysGlySerGluLeuThrCysProGluAspAsnLys 380

QY 1956 ATCAGAGAACTAAGCTTGAATTCGAGAGACTGAAGAAACGGCTCCAGCAAGTTGGAGGTG 2015
Db 381 IleLysGluLeuThrLeuGluIleGluArgLeuLysLysArgLeuGlnGlnLeuGluVal 400

QY 2016 GTGAGGGGGACTTGATGAAGACCGAGGACGGAATATGACCAGTTGGAGCAGAAGTTTCAGA 2075
Db 401 ValGluGlyAspLeuMetLysThrGluAspGluTyrAspGlnLeuGluGlnLysPheArg 420

QY 2076 ACCGAGCAGGATAAGGCAAACTTCCTCTCCAGCAGCTCGAGGAAATCAAAACCAAAATG 2135
Db 421 ThrGluGlnAspLysAlaAsnPheLeuSerGlnGlnLeuGluGluIleLysHisGlnIle 440

QY 2136 GCCAAGCACAAAGCCATAGAGAAAGGGGAGCGCGTGAGCCAGGAAGCCGAACTGCGACAC 2195
Db 441 AlaLysAsnLysAlaIleGluLysGlyGluValValSerGlnGluAlaGluLeuArgHis 460

QY 2196 AGTTTTCGGTGGAGGAGGCTAAAAAGTCGTGATTTACAGGCCGAGGTGAGGCTCTCAAG 2255
Db 461 ArgPheArgLeuGluGluAlaLysSerArgAspLeuLysAlaGlyValGlnAlaLeuLys 480

QY 2256 GAGAAGATCCAGAGCTGATGAACAAGGAAGACCAGCTGTCTCAGTCCAGTCCAGTCACTAT 2315
Db 481 GluLysIleHisGluLeuMetAsnLysGluAspGlnLeuSerGlnLeuGlnValAspTyr 500

QY 2316 TCGTCTCTTCAGCAAAAGATTTATGGAAGAGAAGAACTAAGAACAAGAACATGGGAGGGAG 2375
Db 501 SerValLeuGlnArgPheMetGluGluGluAsnLysAsnLysAsnMetGlyGlnGlu 520

QY 2376 GTCCTCAATCTGACCAAGGAGCTAGAGCTTCCAAAGCGCTACAGCCGAGCTCTCAGGCCG 2435
Db 521 ValLeuAsnLeuThrLysGluLeuLeuSerLysArgTyrSerArgAlaLeuArgPro 540

QY 2436 AGTGGGAACGGCCGAAGGATGGTGGACGTGCTGGCTCCACTCCACTGGGTGCAGACCCGAG 2495
Db 541 SerValAsnGlyLysArgMetValAspValProValThrSerThrGlyValGlnThrAsp 560

QY 2496 GCGGTGTGCGGGGATGCTGCGGAGGAGGAGACCCCGGCTGTGTTTCATTCGCAAAATCCTTC 2555
Db 561 AlaValSerGlyGluAlaAlaGluGluGluThrProAlaValPheIleArgLysSerPhe 580

QY 2556 CAGGAGGAAAATCACATCATGAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCATGGAA 2615
Db 581 GlnGluGluAsnHisIleMetSerAsnLeuArgGlnValGlyLeuLysLysProValGlu 600

QY 2616 CGGTCTCTCGGTCCTCGACAGGTATCCCCCAGCAGCGAAATGAGCTCACCATGAGGAAGTCT 2675
Db 601 ArgSerSerValLeuAspArgTyrProProAlaAlaAsnGluLeuThrMetArgLysSer 620

QY	2676	TGGATTCTCTGGATGAGAAAAAGAGAAAACGGTCTCTTCCACTCCGCAGGAGAAAGGGCCC	2735
Db	621	ProT pMetArgLysArgGluAsnGlyProSerIleThrGlnGluLysGlyPro	640
QY	2736	AGGCCAAACACAGGTGCAGGGCACCCCGGGAGCTGGTCCTAGCACCAAGACGAGGCCAG	2795
Db	641	ArgThrAsnSerSerProGlyHisProGlyGluValValLeuSerProLysGlnGlyGln	660
QY	2796	CCCCTACACATCCGTGTGACACCCAGATCATGAGAACAGCACTGCCACCTGGAGATCACA	2855
Db	661	ProLeuHisIleArgValThrProAspHisGluAsnSerThrAlaThrLeuGluIleThr	680
QY	2856	AGCCCCACATCTGAAGAGTTTTTCTCTAGTACCACCGTCATTCTTACCTTAGGCAACAG	2915
Db	681	SerProThrSerSerGluGluPhePheSerSerThrThrValIleProThrLeuGlyAsnGln	700
QY	2916	AAACCAAGAATAACCATATTTCATCACCCCAATGTCTATGTCGCAAAAGCCCCAAAAGTGCA	2975
Db	701	LysProArgIleThrIleIleProSerProAsnValMetProGlnLysGlnLysSerGly	720
QY	2976	GATCCTACTCTCGGCCCAGAACGAGCCATGTCCCCTGTCAAGATTACTACTATTTCCAGA	3035
Db	721	AspThrThrLeuGlyProGluArgAlaMetSerProValThrIleThrThrPheSerArg	740
QY	3036	GAGAAGAGCCCGAAGGTGGAAGGAGCGCCTTTGCCGACAGGCGCTGCATCCCCCATCAA	3095
Db	741	GluLysThrProGluSerGlyArgGlyAlaPheAlaAspArgProThrSerProIleGln	760
QY	3096	ATCATGACGGTGTCAACATCTGCAGTCCCACTGAAATCGCTGTCTCTCTGAATCTCAG	3155
Db	761	IleMetThrValSerThrSerAlaAlaProAlaGluIleAlaValSerProGluSerGln	780
QY	3156	GAAGTGCCTATGGGAAGACTATCCTCAAAGTCACCCCGGAAAAACAACTGTTCCAGCC	3215
Db	781	GluMetProMetGlyArgThrIleLeuLysValThrProGluLysGlnThrValProThr	800
QY	3216	CCCGTGGGAAGTACAACTCCAATGCTAATATCATCACACCGGAAGACAATAAAATTCAC	3275
Db	801	ProValArgLysTyrAsnSerAsnAlaAsnIleIleThrThrGluAspAsnLysIleHis	820
QY	3276	ATTCACCTGGGTCTCAAGTTTAAGCGATCTCCTGGGCGTCCGCTGAAGCGGTGAGCCCA	3335
Db	821	IleHisLeuGlySerGlnPheLysArgSerProGlyThrSerGlyGluGlyValSerPro	840
QY	3336	GTTATCACCGTCCCGGCTGTCAACGTGACAGCGGAGAGGAGGTTTCTACAGGCACAGTC	3395
Db	841	ValIleThrValArgProValAsnValThrAlaGluLysGluValSerThrGlyThrVal	860
QY	3396	CTTCGCTCTCCCAGGAACCACTCTCTTCAAGACCCGGTCTAGCAAAAGTGACCGCACT	3455
Db	861	LeuArgSerProArgAsnHisLeuSerSerArgProGlyAlaSerLysValThrSerThr	880
QY	3456	ATAACTATAACCCCGGTCAACGTCATCCACACGAGGAACCCCAATCAGTGTCAAGCAA	3515
Db	881	IleThrIleThrProValThrThrSerSerAlaArgGlyThrGlnSerValSerGlyGln	900
QY	3516	GATGGTCACTCTCAGCGGCCTACCCCCACCCGCATTCTTATGTCAAAAGGTATGAAAGCT	3575
Db	901	AspGlySerSerSerGlnArgProThrProThrArgIleProMetSerLysGlyMetLysAla	920
QY	3576	GGAAAGCCAGTAGTGGCAGCCCTCAGGAGCAGGAAATCTGACCAAATTCAGCCTCGAGCT	3635
Db	921	GlyLysProValValAlaAlaProGlyAlaGlyAsnLeuThrLysPheGluProArgAla	940
QY	3636	GAGACTCAGTCTATGAAATATAGACTGAAGAAATCTGCAGCCAGCAGCACTGCCTCTCTT	3695
Db	941	GluThrGlnSerMetLysIleGluLeuLysLysSerAlaAlaSerSerThrThrSerLeu	960
QY	3696	GGAGGGGGGAAGGC	3710
Db	961	GlyGlyGlyLysGly	965

RESULT 7			
Q9NV18			
ID	Q9NV18	PRELIMINARY;	PRT; 653 AA.
AC	Q9NV18;		
DT	01-OCT-2000	(TrEMBLrel. 15, Created)	
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)	
DT	01-OCT-2002	(TrEMBLrel. 22, Last annotation update)	
DE	Hypothetical protein FLJ10708.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Isogai T., Ota T., Hayaashi K., Sugiyama T., Otsuki T., Suzuki Y.,		
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,		
RA	Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Takiguchi S.,		
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,		
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,		
RA	Yanamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,		
RA	Ninomiya K., Iwayanagi T.;		
RT	"NEDO human cDNA sequencing project.";		
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AK001570; BAA91763.1; -.		
KW	Hypothetical protein.		
SQ	SEQUENCE	653 AA; 72451 MW; 33379801AD14AC86 CRC64;	
Alignment Scores:			
Pred. No.:	9.13e-143	Length:	653
Score:	3051.00	Matches:	603
Percent Similarity:	95.25%	Conservative:	19
Best Local Similarity:	92.34%	Mismatches:	31
Query Match:	39.46%	Indels:	0
DB:	4	Gaps:	0
US-10-788-793-1 (1-4364) x Q9NV18 (1-653)			
QY	1752	ATGGAGGAAAAGGAGTACAGTCTGACAAAAGGAGAGGATGAGCTGATGGGTAAACTGAGG	1811
Db	1	MetGluGluLysValTyrAsnLeuThrArgGluArgGluLeuIleGlyLysLeuLys	20
QY	1812	AGCGAAGAAGAAAGGTCTGTGAAGTGAAGTGCAGTGCAGTCTTACTAAAGACGGCTT	1871
Db	21	SerGluGluGluLysSerSerGluLeuSerCysSerValAspLeuLeuLysLysArgLeu	40
QY	1872	GATGSCATAGAGGAGGTAGAAAGGGAATAAACCCGAGGTAGGTCTGCAAGGGGTCTGAG	1931
Db	41	AspGlyIleGluGluValGluArgGluIleThrArgGlyArgSerArgLysGlySerGlu	60
QY	1932	TTCACCTGCCCGGAAGACAATAAGATCAGAGAATTAACGCTTGAATCGAGAGACTGAAG	1991
Db	61	LeuThrCysProGluAspAsnLysIleLysGluLeuThrLeuGluIleGluArgLeuLys	80
QY	1992	AAACGGCTCCAGCAGTTGGAGGTGGTGGAGGGGACTTGATGAAGACCGGACGAATAT	2051
Db	81	LysArgLeuGlnGlnLeuGluValValGluGlyAspLeuMetLysThrGluAspGluTyr	100
QY	2052	GACCAGTTGGAGCAGAAAGTTTCAGAACCGAGCAGGATAAGGCAAACTTCCTCTCCCAGCAG	2111
Db	101	AspGlnLeuGluGlnLysPheArgThrGluGlnAspLysAlaAsnPheLeuSerGlnGln	120
QY	2112	CTCAGGAAATCAAAACACCAAATGGCCAAGCACAAAGCCATAGAGAAAGGGAGGCCGTG	2171
Db	121	LeuGluGluIleLysHisGlnIleAlaLysAsnLysAlaIleGluLysGlyGluValVal	140
QY	2172	AGCCAGGAAGCCGAAGTTCGACACACAGGTTTCGGCTCGAGGAGGCTAAAAGTCGTGATTTA	2231
Db	141	SerGlnGluAlaGluLeuArgHisArgPheArgLeuGluGluAlaLysSerArgAspLeu	160
QY	2232	CAGGCCGAGGTGCAGGCTCTCAAGGAGAAGATCCACGAGCTGATGAACAAGGAAGACCAG	2291
Db	161	LysAlaGluValGlnAlaLeuLysGluLysIleHisGluLeuMetAsnLysGluAspGln	180

QY 2292 CTGTTCTCAGCTCCAAGTCGACTATTTCGGTCTCTCAGCAAGATTTATGGAAGAAGAACT 2351
|||||
Db 181 LeuSerGlnLeuGlnValAspTyrSerValLeuGlnGlnArgPheMetGluGluAsn 200
|||||
QY 2352 AAGAACAAAGAACATGGGAGGAGGTCTCTCAATCTGACCAAGGAGCTAGAGCTTTCCAAAG 2411
|||||
Db 201 LysAsnLysAsnMetGlyGlnGluValLeuAsnLeuThrLysGluLeuLeuSerLys 220
|||||
QY 2412 CGCTACAGCCGAGCTCTCAGGCCGAGTGGAAACGGCCGAGGATGGTGGACGTGCTGTG 2471
|||||
Db 221 ArgTyrSerArgAlaLeuArgProSerValAsnGlyArgArgMetValAspValProVal 240
|||||
QY 2472 GCCTCCACTGGGTGCAGACCGAGGCGGTGTGCGGGGATGCTGCGGAGGAGGAGACCCCG 2531
|||||
Db 241 ThrSerThrGlyValGlnThrAspAlaValSerGlyGluAlaAlaGluGluThrPro 260
|||||
QY 2532 GCTGTGTTTCATTCGCAATCTTCCAGGAGGAAATCACATCATGAGTAATCTTCGACAG 2591
|||||
Db 261 AlaValPheIleArgLysSerPheGlnGluGluAsnHisIleMetSerAsnLeuArgGln 280
|||||
QY 2592 GTAGGCCTGAAGAAACCCATGGAACGGTCTCGTCTCGACAGGTATCCCCCAGCAGCG 2651
|||||
Db 281 ValGlyLeuLysLysProValGluArgSerSerValLeuAspArgTyrProProAlaAla 300
|||||
QY 2652 AATGAGCTCACCATGAGGAAGTCTTGGATTCTTGGATGAGAAAAAGAGAAACGGTCT 2711
|||||
Db 301 AsnGluLeuThrMetArgLysSerTyrIleProTyrMetArgLysArgGluAsnGlyPro 320
|||||
QY 2712 TCCACTCCGAGGAGAAAGGCCCCAGGCCAAACCCAGGGTGCAGGGCACCCCGGGAGCTG 2771
|||||
Db 321 SerIleThrGlnGluLysGlyProArgThrAsnSerSerProGlyHisProGlyGluVal 340
|||||
QY 2772 GTCCTAGCACCAAGCAGGGCCAGGCCCTACACATCCGTGTGACACCAGATCATGAGAAC 2831
|||||
Db 341 ValLeuSerProLysGlnGlyGlnProLeuHisIleArgValThrProAspHisGluAsn 360
|||||
QY 2832 AGCACTGCCACCCCTGGAGATCACAAAGCCCCACATCTGAAGAGTTTTTCTCTAGTACCAC 2891
|||||
Db 361 SerThrAlaThrLeuGluIleThrSerProThrSerGluGluPhePheSerSerThrThr 380
|||||
QY 2892 GTCATTCTACCTTAGGCAACCCAGAAACCAAGATAACCATATTATCCATCACCAATGTC 2951
|||||
Db 381 ValIleProThrLeuGlyAsnGlnLysProArgIleThrIleIleProSerProAsnVal 400
|||||
QY 2952 ATGTCGCAAAAGCCAAAAGTGCAGATCCTACTCTCGGCCCCAGAACGAGCCATGTCCCT 3011
|||||
Db 401 MetProGlnLysGlnLysSerGlyAspThrThrLeuGlyProGluArgAlaMetSerPro 420
|||||
QY 3012 GTCACGATTACTACTATTTCAGAGAGAAGAGCCCGGAAGGTGGAAGGAGCGCCTTTGCC 3071
|||||
Db 421 ValThrIleThrThrPheSerArgGluLysThrProGluSerGlyArgGlyAlaPheAla 440
|||||
QY 3072 GACAGGCCTGCATCCCCCATCCAAATCATGACGGTGTCAACATCTGCAGCTCCCACTGAA 3131
|||||
Db 441 AspArgProThrSerProIleGlnIleMetThrValSerThrSerAlaAlaProAlaGlu 460
|||||
QY 3132 ATCGCTGTCTCTCTGAATCTCAGGAAGTGCCTATGGGAAGGACTATCCTCAAAGTCAACC 3191
|||||
Db 461 IleAlaValSerProGluSerGlnGluMetProMetGlyArgThrIleLeuLysValThr 480
|||||
QY 3192 CCGGAAAAACAACTGTTCCAGCCCCCGTGCAGGAGTACAACTCCAATGCTAATATCATC 3251
|||||
Db 481 ProGluLysGlnThrValProThrProValArgLysTyrAsnSerAsnAlaAsnIleIle 500
|||||
QY 3252 ACCACGGAAGACAATAAATTCACATTCACCTGGGTCTCAGTTTAAGCGATCTCCTGGG 3311
|||||
Db 501 ThrThrGluAspAsnLysIleHisIleHisLeuGlySerGlnPheLysArgSerProGly 520
|||||
QY 3312 CCTGCCGTGAAGCGTGAGCCCCAGTTATCACCGTCCGGCCTGTCAACGTGACAGCGGAG 3371
|||||
Db 521 ThrSerGlyGluGlyValSerProValIleThrValArgProValAsnValThrAlaGlu 540
|||||
QY 3372 AAGGAGGTTTCTACAGGCACAGTCTCTCGCTCTCTCCAGGAACCACTCTCTTCAAGACCC 3431
|||||

Db 541 LysGluValSerThrGlyThrValLeuArgSerProArgAsnHisLeuSerSerArgPro 560
|||||
QY 3432 GGTGTAGCAAAAGTGACCAAGCAGCACTATAACTATAACCCCGGTCAACAGTCATCCACAGA 3491
|||||
Db 561 GlyAlaSerLysValThrSerThrIleThrIleThrProValThrThrSerSerAlaArg 580
|||||
QY 3492 GGAACCCAATCAGTGTCTCAGGACAAGATGGGTCTCTCAGCGGCCTACCCCAACCCGCAAT 3551
|||||
Db 581 GlyThrGlnSerValSerGlyGlnAspGlySerSerGlnArgProThrProThrArgIle 600
|||||
QY 3552 CCTATGTCAAAGGTATGAAAGCTGGAAGCCAGTGTGGCAGCTCTCAGGAGCAGGAAAT 3611
|||||
Db 601 ProMetSerLysGlyMetLysAlaGlyLysProValValAlaAlaProGlyAlaGlyAsn 620
|||||
QY 3612 CTGACCAAAATTCAGCCTCGAGCTCGAGCTGAGACTCAGTCTATGAAAATAGAGCTGAAGAAATCT 3671
|||||
Db 621 LeuThrLysPheGluProArgAlaGluThrGlnSerMetLysIleGluLeuLysLysSer 640
|||||
QY 3672 GCAGCCAGCAGCACTGCCTCTCTTGGAGGGGGAAGGGC 3710
|||||
Db 641 AlaAlaSerSerThrThrSerLeuGlyGlyGlyLysGly 653
|||||
RESULT 8
Q9CS72
ID Q9CS72 PRELIMINARY; PRT; 610 AA.
AC Q9CS72;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE 5730485H21Rik protein (Fragment).
GN 5730485H21RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK017709; BAB30888.1; -;
DR MGI; MGI:1917848; 5730485H21Rik.
DR InterPro; IPR000533; Tropomyosin.
DR PRINTS; PR00194; TROPOMYOSIN.
FT NON_TER 610 610
SQ SEQUENCE 610 AA; 70346 MW; 9B9EFCB7F64FCE80 CRC64;

Alignment Scores:
Pred. No.: 1.97e-117 Length: 610
Score: 2537.50 Matches: 509
Percent Similarity: 99.23% Conservative: 6
Best Local Similarity: 98.07% Mismatches: 3
Query Match: 32.82% Indels: 1

DB:	11	Gaps:	1
US-10-788-793-1 (1-4364) x Q9CS72 (1-610)			
QY	66	CAGGTGGGAATGAGATCACGAAATCAAGGTGGAGAAAAGTTTCATCTTAACCGGCATGTCCTC	125
Db	92	GlnValGlyMetArgSerArgAsnGlnGlyGlySerSerAsnGlyHisValSer	111
QY	126	TGCCCCAAAGTCCTCCATCATCAGCAGTGATGGTGGTAAGGCCCTTCAGAAAGATGCAAAA	185
Db	112	CysProLysProSerIleIleSerSerAspGlyGlyLysGlyProSerGluAspAlaLys	131
QY	186	AAGAACAAAGCCAATCGGRAG---GAGGAGGATGTTCATGGCTTCGGGAACCTATCAAAAAGG	242
Db	132	LysAsnLysAlaAsnArgLysGlyGluAspValMetAlaSerGlyThrValLysArg	151
QY	243	CACCTCAACCATCTGGAGAAAAGTGAGAAAAAGACTAAGAAGTCTGTGGAGTTATCCAAAG	302
Db	152	HisLeuLysProSerGlySerGluLysLysThrLysLysProLeuGluLeuSerLys	171
QY	303	GAGGACCTCATCCAGCTCTGTAGTATCATTGGAAAGGGAGTTGCAGGCTCGAGAAGATGTC	362
Db	172	GluAspLeuIleGlnLeuLeuSerIleMetGluGlyGluLeuGlnAlaArgGluAspVal	191
QY	363	ATCCACATGCTGAGGACAGAGAAAACCAAGCCCCGAGGTTCTGGAGGCCACACTATGGATCT	422
Db	192	IleHisMetLeuLysThrGluLysThrLysProGluValLeuGluAlaHisTyrGlySer	211
QY	423	GCAGAACTGAGAAAGTCTTCGGGTCTCGGCACCGAGATGCCATCCTTGCTCAAGAGAAG	482
Db	212	AlaGluProGluLysValLeuArgValLeuHisArgAspAlaIleLeuAlaGlnGluLys	231
QY	483	TCCATAGGAGAAGACGTTCTATGAGAAAACCTATCTCAGAGCTGGACAGACTGGAGGAAAAG	542
Db	232	SerIleGlyGluAspValTyrGluLysProIleSerGluLeuAspArgLeuGluGluLys	251
QY	543	CAGAAGGAGACGTACCGCGCATGCTAGACGAGCTGCTGCTGGCTGAGAAGTGTCCACAGG	602
Db	252	GlnLysGluThrTyrArgArgMetLeuGluGlnLeuLeuAlaGluLysCysHisArg	271
QY	603	CGCACCGTGTCGAGCTGGAGAACGAGAAGCACAAGCACACTGACTACATGAAACAAGAGC	662
Db	272	ArgThrValTyrGluLeuGluAsnGluLysHisLysHisThrAspTyrMetAsnLysSer	291
QY	663	GACGACTTCACCAACCTGCTGGAGCGAGGAGAGAGGTTGAAAAAGCTCCTTGAAACAA	722
Db	292	AspAspPheThrAsnLeuLeuGluGlnGluArgGluArgLeuLysLysLeuLeuGluGln	311
QY	723	GAAAAAGCTTACCAAGCCGCAAGAAAAGGAAAACGCTAAGCGGCTCAACAACTCGA	782
Db	312	GluLysAlaTyrGlnAlaArgLysGluLysGluAsnAlaLysArgLeuAsnLysLeuArg	331
QY	783	GATGAGCTTGTGAAGCTCAAGTCCCTTCGCCCTCATGTTGGTGGACGAGAGGCAGATGCAC	842
Db	332	AspGluLeuValLysLeuLysSerPheAlaLeuMetLeuValAspGluArgGlnMetHis	351
QY	843	ATCAGCAACTGGGCCTGCAGAGTCAAGAAAGTCCAGGACCTCACTCAGAAAGCTGAGGGAG	902
Db	352	IleGluGlnLeuGlyLeuGlnSerGlnLysValGlnAspLeuThrGlnLysLeuArgGlu	371
QY	903	GAGGAAGAAAACCTCAAAGCGGTCACTTACAAATCCAAGGAAGACCGCCAGAGCTGCTC	962
Db	372	GluGluGluLysLeuLysAlaIleThrTyrLysSerLysGluAspArgGlnLysLeuLeu	391
QY	963	AAGTTAGAAGTGGACTTCGAACACAAAGGCCTCGAGGTTTCCAGGAGCACGAAGAGATG	1022
Db	392	LysLeuGluValAspPheGluHisLysAlaSerArgPheSerGlnGluHisGluGluMet	411
QY	1023	AACGCCAAATTGGCGGAATCAAGAATCTCACAAACCGGCAACTTCGACTCAAACTGGTGGC	1082
Db	412	AsnAlaLysLeuAlaAsnGlnGluSerHisAsnArgGlnLeuArgLeuLysLeuValGly	431
QY	1083	TTATCGCAAAGGATTGAGGAGCTGGAAGAGAGACCAATAAAAAAGCCTTCAGAAGGCAGAGAA	1142

Db	432	LeuSerGlnArgIleGluGluLeuGluGluThrAsnLysSerLeuGlnLysAlaGluGlu	451
QY	1143	GAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAACCTCCAGTCTCATG	1202
Db	452	GluLeuGlnGluLeuArgAspLysIleAlaLysGlyGluCysGlyAsnSerSerLeuMet	471
QY	1203	GCGGAAGTGGAGAGTCTGCGCAAGCGCTGCTTGAGATGGAGGGCAAGGATGAAGAGATC	1262
Db	472	AlaGluValGluSerLeuArgLysArgValLeuGluMetGluGlyLysAspGluGluIle	491
QY	1263	ACGAAGACCGAGGCGCCAGTCCCGGAGCTGAAGAAGAGCTCCAAGAGGAAGAACCCAC	1322
Db	492	ThrLysThrGluAlaGlnCysArgGluLeuLysLysLysLeuGlnGluGluHisHis	511
QY	1323	AGCAAGAACTTAGACTAGAAAGTGGAGAAGCTGCAGAAGGATGTCTGAGCTGGAGAAG	1382
Db	512	SerLysGluLeuArgLeuGluValGluLysLeuGlnLysArgMetSerGluLeuGluLys	531
QY	1383	CTGGAGGAAGCGTTTCAGCCGGAGTAAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAG	1442
Db	532	LeuGluGluAlaPheSerArgSerLysSerGluCysThrGlnLeuHisLeuAsnLeuGlu	551
QY	1443	AAGGAGAAAGAACCTTAACCAAGACCTGCTGTAACGAGCTGGAGGTGTCAAAGAGTCGAGTT	1502
Db	552	LysGluLysAsnLeuThrLysAspLeuLeuAsnGluLeuGluValLysSerArgVal	571
QY	1503	AAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAGCGCGAGTTAAGCCTCAAAGATGAC	1562
Db	572	LysGluLeuGluCysSerGluSerArgLeuGluLysValGluLeuSerLeuLysAspAsp	591
QY	1563	CTTACAAAAGCTGAAGTCCTTCACGTGTGATGCTGGTGGATGAGAGGAAAAATATGATG	1619
Db	592	LeuThrLysLeuLysSerPheThrValMetLeuValAspGluArgLysAsnMetMet	610
RESULT 9			
Q8N8B9	PRELIMINARY;	PRT;	512 AA.
ID	Q8N8B9		
AC	Q8N8B9;		
DT	01-OCT-2002 (TrEMBLrel. 22, Created)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	Hypothetical protein FLJ39702 (Fragment).		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Small intestine;		
RA	Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,		
RA	Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,		
RA	Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,		
RA	Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,		
RA	Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,		
RA	Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,		
RA	Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,		
RA	Isogai T.;		
RT	"NEDO human cDNA sequencing project.";		
RL	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AK097021; BAC04928.1; -.		
DR	InterPro; IPR000533; Tropomyosin.		
DR	PRINTS; PR00194; TROPOMYOSIN.		
KW	Hypothetical protein.		
FT	NON TER 512		
SQ	SEQUENCE 512 AA; 60035 MW; 41F3B8A2FBD28E20 CRC64;		
Alignment Scores:			
Pred. No.:	5.66e-111	Length:	512
Score:	2406.50	Matches:	482
Percent Similarity:	97.66%	Conservative:	18
Best Local Similarity:	94.14%	Mismatches:	11
Query Match:	31.13%	Indels:	1

DB:	4	Gaps:	1
US-10-788-793-1 (1-4364) x Q8N8B9 (1-512)			
QY	75	ATGAGATCAGCAATCAAGTGGAGAAAGTTTCATCTAACGGGCATGTCTCTCGCCCAAG	134
Db	1	MetArgSerArgAsnGlnGlyGluSerAlaSerAspGlyHisIleSerCysProLys	20
QY	135	TCCTCCATCATCAGCAGTGGTGGTAAGGGCCCCCTCAGAAGATGCA--	191
Db	21	ProSerIleIleGlyAsnAlaGlyGluLysSerLeuSerGluAspAlaLysLysLys	40
QY	192	AAGGCCAATCGGAAGGAGGAGTGTATGGCTTCCGGAACCTATCAAAAGGCACCTCAAA	251
Db	41	LysSerAsnArgLysGluAspValMetAlaSerGlyThrValLysArgHisLeuLys	60
QY	252	CCATCTGGAGAAAGTGAGAAAAGACTAAGAAGTCTGTGGAGTTATCCAAGGAGGACCTC	311
Db	61	ThrSerGlyGluCysGluArgLysThrLysLysSerLeuGluLeuSerLysGluAspLeu	80
QY	312	ATCCAGCTCCTGAGTATCATGGAAGGGGAGTTGCAGGCTCGAGAAGATGTCTATCCACATG	371
Db	81	IleGlnLeuLeuSerIleMetGluGlyGluLeuGlnAlaArgGluAspValIleHisMet	100
QY	372	CTGAGGACAGAGAAAACCAAGCCCGAGGTTCTGGAGGCACACTATGGATCTGCAGAACCT	431
Db	101	LeuLysThrGluLysThrLysProGluValLeuGluAlaHisTyrGlySerAlaGluPro	120
QY	432	GAGAAAGTGCTTCGGGTCTCGACCGGAGATGCCATCCTTGCTCAAGAGAAGTCCATAGGA	491
Db	121	GluLysValLeuArgValLeuHisArgAspAlaIleLeuAlaGlnGluLysSerIleGly	140
QY	492	GAAGACGTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAAAAGCAGAGGAG	551
Db	141	GluAspValTyrGluLysProIleSerGluLeuAspArgLeuGluLysGlnLysGlu	160
QY	552	ACGTACCGCCGCATGCTAGACAGACTGCTGCTGGCTGAGAAAGTGTACAGGCGCACCGTG	611
Db	161	ThrTyrArgArgMetLeuGluGlnLeuLeuLeuAlaGluLysCysHisArgArgThrVal	180
QY	612	TACGAGCTGGAGAACGAGAGCACAAAGCACACTGACTACATGAACAAGAGCGAGCTTC	671
Db	181	TyrGluLeuGluAsnGluLysHisLysHisThrAspTyrMetAsnLysSerAspAspPhe	200
QY	672	ACCAACCTGCTGGAGCAGGAGCGAGAGAGGTTTGAANAAGCTCCTTGAACAAGAAAAGCT	731
Db	201	ThrAsnLeuLeuGluGlnGluArgGluArgLeuLysLysLeuLeuGluGlnGluLysAla	220
QY	732	TACCAAGCCCGCAAGAAAAGGAAAACGCTAAGCGGCTCAACAAACTTCGAGATGAGCTT	791
Db	221	TyrGlnAlaArgLysGluLysGluAsnAlaLysArgLeuAsnLysLeuArgAspGluLeu	240
QY	792	GTGAAGCTCAAGTCTTCGCCCTCATGTTGGTGGACGAGAGGCAGATGCACATCGAGCAA	851
Db	241	ValLysLeuLysSerPheAlaLeuMetLeuValAspGluArgGlnMetHisIleGluGln	260
QY	852	CTGGGCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAAGCTGAGGGAGGAGGAAGAA	911
Db	261	LeuGlyLeuGlnSerGlnLysValGlnAspLeuThrGlnLysLeuArgGluGluGluGlu	280
QY	912	AAACTCAAAGCGGTCACTTACAAATCCAAGGAAGACCGCCAGAAGCTGCTCAAGTTAGAA	971
Db	281	LysLeuLysAlaIleThrSerLysSerLysGluAspArgGlnLysLeuLeuLysLeuGlu	300
QY	972	GTGGACTTCGAACACAAAGSCCTCGAGGTTTTCACGAGGACGACGAAGAGATGAACGCCAA	1031
Db	301	ValAspPheGluHisLysAlaSerArgPheSerGlnGluHisGluGluMetAsnAlaLys	320
QY	1032	TTGGCGAATCAAGAATCTCACAAACCGGCAACTTCGACTCAAACCTGGTTGGCTTATCGCAA	1091
Db	321	LeuAlaAsnGlnGluSerHisAsnArgGlnLeuArgLeuLysLeuValGlyLeuThrGln	340
QY	1092	AGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAGGCAGAGGAGAGCTCCAG	1151

Db	341	ArgIleGluGluLeuGluGluThrAsnLysAsnLeuGlnLysAlaGluGluGluGln	360
QY	1152	GAGCTGAGAGAGAAAATTGCCAAAGGGGAATGTGGAAACTCCAGTCTCATGGCGGAAGTG	1211
Db	361	GluLeuArgAspLysIleAlaLysGlyGluCysGlyAsnSerSerLeuMetAlaGluVal	380
QY	1212	GAGAGTCTGCGCAAGCCGCTGCTTGAGATGGAGGGCAAGGATGAAGAGATCACGAAGACC	1271
Db	381	GluAsnLeuArgLysArgValLeuGluMetGluGlyLysAspGluGluIleThrLysThr	400
QY	1272	GAGGCCAGTCCCGGAGCTGAAGAAGAGCTCCAAGAGGAAGAACACCACAGCAAGGAA	1331
Db	401	GluSerGlnCysArgGluLeuArgLysLysLeuGlnGluGluGluHisHisSerLysGlu	420
QY	1332	CTTAGACTAGAAAGTGAGAAAGCTGCAGAAAGAGGATGTCTGAGCTGGAGAAAGCTGGAGAA	1391
Db	421	PheArgLeuGluValGluLysLeuGlnLysArgMetSerGluLeuGluLysLeuGluGlu	440
QY	1392	GCGTTTCAGCCGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAGGAGAAG	1451
Db	441	AlaPheSerLysSerLysSerGluCysThrGlnLeuHisLeuAsnLeuGluLysGluLys	460
QY	1452	AACCTAACCAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTC	1511
Db	461	AsnLeuThrLysAspLeuLeuAsnGluLeuGluValValLysSerArgValLysGluLeu	480
QY	1512	GAATGCTCCGAGAGTAGACTGGAGAAGGCCGAGTTAAGCCTCAAAGATGACCTTACAAAG	1571
Db	481	GluCysSerGluSerArgLeuGluLysAlaGluLeuSerLeuLysAspLeuThrLys	500
QY	1572	CTGAAGTCTTCACTGTGATGCTGGTGATGAGAGG	1607
Db	501	LeuLysSerPheThrValMetLeuValAspGluArg	512
RESULT 10			
Q8N6Z0			
ID	Q8N6Z0	PRELIMINARY;	PRT; 893 AA.
AC	Q8N6Z0;		
DT	01-OCT-2002	(TReMBLrel. 22, Created)	
DT	01-OCT-2002	(TReMBLrel. 22, Last sequence update)	
DT	01-OCT-2002	(TReMBLrel. 22, Last annotation update)	
DE	Similar to downregulated in ovarian cancer 1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Pancreas;		
RA	Strausberg R.;		
RL	Submitted (APR-2002) to the EMBL/GenBank/DBSJ databases.		
DR	EMBL; BC027860; AAH27860.1; -.		
SQ	SEQUENCE 893 AA; 102232 MW; 561BE17D6C10FDC4 CRC64;		
Alignment Scores:			
Pred. No.:	3.04e-87	Length:	893
Score:	1926.00	Matches:	424
Percent Similarity:	64.21%	Conservative:	168
Best Local Similarity:	45.99%	Mismatches:	266
Query Match:	24.91%	Indels:	64
DB:	4	Gaps:	14
US-10-788-793-1 (1-4364) x Q8N6Z0 (1-893)			
QY	816	ATGTTGGTGGACGAGAGGCAGATGCACATCGAGCAAACTGGGCCTGCAGAGTCAGAAAGTC	875
Db	1	MetValValAspGluGlnGlnArgLeuThrAlaGlnLeuThrLeuGlnArgGlnLysIle	20
QY	876	CAGGACCTCACTCAGAAGCTGAGGGAGGAGGAAGAAAACCTCAAAGCGGTCACTTACAA	935
Db	21	GlnGluLeuThrThrAsnAlaLysGluThrHisThrLysLeuAlaLeuAlaGluAlaArg	40


```
Db 737 ThrMetAlaThrPheAlaArgAlaGlnThrProGluSerCysGlySerLeuThrProGlu 756
QY 3075 AGGCTGCATCCCCCATCCAAATCATGACGGTGTCAACATCTGCAGCTCCCACTGAAATC 3134
Db 757 ArgThrMetSerProIleGlnValLeuAlaValThrGlySerAlaSerSerProGluGln 776
QY 3135 GCTGTCTCTCCTGAATCTCAGGAAGTGCCTATGGGAAGGACTATCTCTCAAAGTCACCCCG 3194
Db 777 GlyArgSerProGluProThrGluIleSerAlaLysHisAlaIlePheArgValSerPro 796
QY 3195 GAAAAACAAACTGTTCCAGCCCCCGTGCAGGAAGTACAACTCCAATGCTAATATCATCACC 3254
Db 797 AspArgGlnSerSerTrpGlnPheGlnArgSerAsnSerAsnSerSerSerValIleThr 816
QY 3255 ACAGAAGACAAATAAAATTCACATTACCTGGGTTCTCAGTTTAAGCGATCTCTCTGGCCCT 3314
Db 817 ThrGluAspAsnLysIleHisIleHisLeuGlySerProTyrMetGlnAla----- 833
QY 3315 GCCGCTGAAGCGCTGAGCCAGTTATCACCCGTCGCCGCTGTCAACGTGACAGCGGAGAAG 3374
Db 834 ValAlaSerProValArgProAlaSerProSerAlaProLeuGlnAspAsnArgThrGln 853
QY 3375 GAGGTTTCTACAGGCACAGTCCTTCGCTCTCCAGGAACCACTCTCTTCAAGACCCGGT 3434
Db 854 GlyLeuIleAsnGlyAlaLeuAsnLysThr----- 863
QY 3435 GCTAGCAAAGTGACCACTATAACTATAACCCCGGTCAACACGTATCCACACGAGGA 3494
Db 864 ThrAsnLysValThrSerSerIleThrIleThrProThrAlaThrProLeuProArgGln 883
QY 3495 ACCCAA 3500
Db 884 SerGln 885
```

RESULT 11

```
Q8IUM3
ID Q8IUM3 PRELIMINARY; PRT; 764 AA.
AC Q8IUM3;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE GPBP-interacting protein 90.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RA Revert-Ros F., Raya A., Granero F., Saus J.;
RT "Cloning and expression of GIP90, a protein down-regulated in cancer
RT cells which interacts with GPBP.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF329092; AAN16206.1; -.
DR InterPro; IPR000533; Tropomyosin.
DR PRINTS; PR00194; TROPOMYOSIN.
SQ SEQUENCE 764 AA; 89785 MW; DB41B4CB4EEFA085 CRC64;
```

Alignment Scores:

Pred. No.:	2.06e-82	Length:	764
Score:	1828.00	Matches:	386
Percent Similarity:	69.55%	Conservative:	144
Best Local Similarity:	50.66%	Mismatches:	210
Query Match:	23.65%	Indels:	22
DB:	4	Gaps:	7

US-10-788-793-1 (1-4364) x Q8IUM3 (1-764)

```
QY 75 ATGAGATCAGCAATCAAGGTGGAGAAAGTTTCATCTAACGGGCATGTCTCTGCCCAAG 134
Db 1 MetArgSerArg-----GlySerAspThrGluGlySerAlaGlnLysLysPheProArg 18
QY 135 TCCTCCATCATCAGCAGTGATGGTGAAGGGCCCTCAGAAGATGCAAAAAAGAACAAAG 194
```

```
Db 19 HisThr-----LysGlyHisSerPheGlnGlyProLysAsnMet 31
QY 195 GCCAATCGGAAGGAGGAG--GATGTCTATGGCTTCGGAAGCTATCAAAAGGCACCTCAAA 251
Db 32 LysHisArgGlnGlnAspLysAspSerProSerGluSerAspVal-----IleLeu 48
QY 252 CCATCTGGAGAAAGTCAGAAA-----AAGACTAAGAAGTCTGTGGAGTTATCC 299
Db 49 ProCysProLysAlaGluLysProHisSerGlyAsnGlyHisGlnAlaGluAspLeuSer 68
QY 300 AAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGGAGTTGCAGGGCTCGAGAAGAT 359
Db 69 ArgAspAspLeuLeuPheLeuLeuSerIleLeuGluGlyGluLeuGlnAlaArgAspGlu 88
QY 360 GTCATCCACATGCTGAGGACAGAGAAAAACCAAGCCCGAGGTTCTGGAGGCACACTATGGA 419
Db 89 ValIleGlyIleLeuLysAlaGluLysMetAspLeuAlaLeuLeuGluAlaGlnTyrGly 108
QY 420 TCTGCAGAACCTGAGAAAAGTCTTCGGGTCTTCGACCCGAGATGCCATCTTGTCTCAAGAG 479
Db 109 PheValThrProLysLysValLeuGluAlaLeuGlnArgAspAlaPheGlnAlaLysSer 128
QY 480 AAGTCCATAGGAGAAAGACGCTCTATGAGAAAACCTATCTCAGAGCTGGACAGACTGGAGGAA 539
Db 129 ThrProTrpGlnGluAspIleTyrGluLysProMetAsnGluLeuAspLysValValGlu 148
QY 540 AAGCAGAAGGAGACCTACCGCCGATGCTAGACGAGCTGCTGGCTGAGAAAGTGTCTAC 599
Db 149 LysHisLysGluSerTyrArgArgIleLeuGlyGlnLeuValAlaGluLysSerArg 168
QY 600 AGGCGCACCGTGTACGAGCTGGAGAACGAGAACACAGCACACTGACTACATCAACAAG 659
Db 169 ArgGlnThrIleLeuGluLeuGluGluLysArgLysHisLysGluTyrMetGluLys 188
QY 660 AGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGAGAGGTTGAAAAAGCTCCTTGAA 719
Db 189 SerAspGluPheIleCysLeuLeuGluGlnGluCysGluArgLeuLysLysLeuIleAsp 208
QY 720 CAAGAAAAAGCTTACCAAGCCCGCAAGAAAAAGAAAAACGCTAAGCGGCTCAACAACTT 779
Db 209 GlnGluIleLysSerGlnGluGluLysGluGlnGluLysGluLysArgValThrThrLeu 228
QY 780 CGAGATGAGCTTGTGAAGCTCAAGTCCCTTCGCCCTCATGTTGGTGGACGAGAGCGCAGATG 839
Db 229 LysGluGluLeuThrLysLeuLysSerPheAlaLeuMetValValAspGluGlnArg 248
QY 840 CACATCGAGCAACTGGGCCCTGCAGAGTCAAGAGTCCAGGACCTCACTCAGAAGCTGAGG 899
Db 249 LeuThrAlaGlnLeuThrLeuGlnArgGlnLysIleGlnGluLeuThrThrAsnAlaLys 268
QY 900 GAGGAGGAAGAAAACTCAAAGCGGTCACTTACAAAATCCAAGGAAGACCCGCCAGAAAGCTG 959
Db 269 GluThrHisThrLysLeuAlaLeuAlaGluAlaArgValGlnGluGluGlnLysAla 288
QY 960 CTCAAAGTTAGAGTGAGCTTCGAACACAAAGGCTCGAGGTTTCCAGGAGCAGCAGAGAG 1019
Db 289 ThrArgLeuGluLysGluLeuGlnThrGlnThrThrLysPheHisGlnAspGlnAspThr 308
QY 1020 ATGAACGCCAAATTGGCGAATCAAGAATCTCACAAACCGGCAACTTCGACTCAAACTGGTT 1079
Db 309 IleMetAlaLysLeuThrAsnGluAspSerGlnAsnArgGlnLeuGlnGlnLysLeuAla 328
QY 1080 GGCTTATCGCAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAGGCAGAG 1139
Db 329 AlaLeuSerArgGlnIleAspGluLeuGluGluThrAsnArgSerLeuArgLysAlaGlu 348
QY 1140 GAAGAGCTCCAGGAGCTGAGAGAGAAAAATTCGCAAAAGGGGAATGTGGAACCTCCAGTCTC 1199
Db 349 GluGluLeuGlnAspIleLysGluLysIleSerLysGlyGluTyrGlyAsnAlaGlyIle 368
QY 1200 ATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTCTTGAGATGGAGGGCAAGGATGAAGAG 1259
```

Db 369 MetAlaGluValGluGluLeuArgLysArgValLeuAspMetGluGlyLysAspGluGlu 388
QY 1260 ATCAGGAAGACCGAGGCCCACTGCCGGAGCTGAAGAAGAGCTCCAAGAGGAACAAC 1319
Db 389 LeuileLysMetGluGluGlnCysArgAspLeuAsnLysArgLeuGluArgGluThrLeu 408
QY 1320 CACAGCAAGGAATTAGACTAGAAAGTGGAGAAGCTGCAGAAAGGAGGTGCTGAGCTGGAG 1379
Db 409 GlnSerLysAspPheLysLeuGluValGluLysLeuSerLysArgIleMetAlaLeuGlu 428
QY 1380 AAGCTGGAGGAAGCGTTACGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTG 1439
Db 429 LysLeuGluAspAlaPheAsnLysSerLysGlnGluCysTyrSerLeuLysCysAsnLeu 448
QY 1440 GAGAAGGAGAAGAACCTTAACCAAGACCTGCTGAACGAGCTGGAGGTGCTCAAGAGTCGA 1499
Db 449 GluLysGluArgMetThrThrLysGlnLeuSerGlnGluLeuGluSerLeuLysValArg 468
QY 1500 GTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAGCCGAGTTAAGCCTCAAGAT 1559
Db 469 IleLysGluLeuGluAlaIleGluSerArgLeuGluLysThrGluPheThrLeuLysGlu 488
QY 1560 GACCTTACAAAAGCTGAAGTCCTTCACCTGTGATGCTGGTGGATGAGAGGAAAAATATGATG 1619
Db 489 AspLeuThrLysLeuLysThrLeuThrValMetPheValAspGluArgLysThrMetSer 508
QY 1620 GAGAAATAAAGCAAGAGAGAGGAAAGTGGATGGTTCAATAAAAACTTTAAGTGGAG 1679
Db 509 GluLysLeuLysLysThrGluAspLysLeuGlnAlaIaSerSerGlnLeuGlnValGlu 528
QY 1680 CAGGGAAGAGTCATGGATGTGACGGAAAAAGCTAATCGAGGAAAGCAAGAAGCTTTTAAAA 1739
Db 529 GlnAsnLysValThrThrValThrGluLysLeuIleGluLysThrLysArgAlaLeuLys 548
QY 1740 CTCAAATCTGAAATGGAGGAAAAAGGACTACAGTCTGACAAAGGAGAGGATGAGCTGATG 1799
Db 549 SerLysThrAspValGluGluLysMetTyrSerValThrLysGluArgAspAspLeuLys 568
QY 1800 GGTAACCTGAGGACCGAAGAAAGGTCTGTGAACTGAGCTGCAGTGTAGACTTACTA 1859
Db 569 AsnLysLeuLysAlaGluGluLysGlyAsnAspLeuLeuSerArgValAsnMetLeu 588
QY 1860 AAGAAGCGCTTGATGGCATAGAGGAGGTAGAGGGGAAATAAACCCGAGGTAGG----- 1913
Db 589 LysAsnArgLeuGlnSerLeuGluAlaIleGluLysAspPheLeuLysAsnLysLeuAsn 608
QY 1914 -----TCGTGCAAGGGTCTGAGTTACCTGCTCCCGAAGACAATAAGATCAGAGACTA 1967
Db 609 GlnAspSerGlyLysSerThrThrAlaLeuHisGlnGluAsnAsnLysIleLysGluLeu 628
QY 1968 ACGCTTGAAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTGGTGGAGGGGAC 2027
Db 629 SerGlnGluValGluArgLeuLysLeuLysLeuLysAspMetLysAlaIleGluAspAsp 648
QY 2028 TTGATGAAGACCGAGGACGAATATGACCAGTTGGAGCAGAAGTTTCAGAAACCGAGCAGGAT 2087
Db 649 LeuMetLysThrGluAspGluTyrGluThrLeuGluArgArgTyrAlaAsnGluArgAsp 668
QY 2088 AAGGCATACTTCCTCTCCAGCAGCTCGAGGAAATCAAAACCAATGGCCCAAGCAAAA 2147
Db 669 LysAlaGlnPheLeuSerLysGluLeuGluHisValLysMetGluLeuAlaLysTyrLys 688
QY 2148 GCCATAGAGAAAGGGGAGGCCGTGAGCCAGGAAGCCGAACCTCGACACAGGTTTCGCGCTG 2207
Db 689 LeuAlaGluLysThrGlu---ThrSerHisGlnGlnTrpLeuPheLysArgLeuGlnGlu 707
QY 2208 GAGGAGGCTAAAAGTCGTGATTTACAGGCCGAGGTGCGGCTCTCAAGGAGAAGATCCAC 2267
Db 708 GluGluAlaLysSerGlyHisLeuSerArgGluValAspAlaLeuLysGluLysIleHis 727
QY 2268 GAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTATTTCGTCCTTCAG 2327
Db 728 GluTyrMetAlaThrGluAspLeuIleCysHisLeuGlnGlyAspHisSerValLeuGln 747

QY 2328 CAAAGA 2333
Db 748 LysLys 749
RESULT 12
Q13597
ID Q13597 PRELIMINARY; PRT; 752 AA.
AC Q13597;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE DOC1.
GN DOC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP -SEQUENCE FROM N.A.
RX MEDLINE=94148289; PubMed=8314147;
RA Mok S.C., Wong K.K., Chan R.K., Lau C.C., Tsao S.W., Knapp R.C.,
RA Berkowitz R.S.;
RT "Molecular cloning of differentially expressed genes in human
epithelial ovarian cancer.";
RL Gynecol. Oncol. 52:247-252(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Wong K.K., Mok S.C.;
RT "Cloning and sequencing of full length Doc1 and Doc2 mRNAs.";
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53445; AAA98972.1; -.
DR GO; GO:0016459; C:myosin; NAS.
SQ SEQUENCE 752 AA; 87141 MW; D763B4E2806A3B0B CRC64;

Alignment Scores:
Pred. No.: 2.08e-73 Length: 752
Score: 1645.50 Matches: 363
Percent Similarity: 64.16% Conservative: 140
Best Local Similarity: 46.30% Mismatches: 216
Query Match: 21.28% Indels: 65
DB: 4 Gaps: 13

US-10-788-793-1 (1-4364) x Q13597 (1-752)

QY 816 ATGTTGGTGACGAGAGGCGAGCATCGAGCAACTGGGCCCTGCAGAGTCAGAAAGTC 875
Db 1 MetValValAspGluGlnGlnArgLeuThrAlaGlnLeuThrLeuGlnArgGlnLysIle 20
QY 876 CAGGACCTCACTCAGAAAGCTGAGGAGGAGGAGGAGAAAACTCAAAAGCGGTCACTTACAAA 935
Db 21 GlnGluLeuThrThrAsnAlaLysGluThrHisThrLysLeuAlaLeuAlaGluAlaArg 40
QY 936 TCCAAGGAAGACCGCCAGAAGCTGCTCAAGTTAGAAGTGGACTTCGAAACACAAGGCTCG 995
Db 41 ValGlnGluGluGlnLysAlaThrArgLeuGluLysGluLeuGlnThrGlnThrThr 60
QY 996 AGGTTTTCCTCCAGGAGCACGAAGAGATGAACGCCAAATTGGCGAATCAAGAAATCTCACAAC 1055
Db 61 LysPheHisGlnAspGlnAspThrIleMetAlaLysLeuThrAsnGluAspSerGlnAsn 80
QY 1056 CGGCAACTTCGACTCAAACTGGTTGGCTTATCGCAAAAGGATTGAGGAGCTGGAAGAGACC 1115
Db 81 ArgGlnLeuGlnGlnLysLeuAlaLeuSerArgGlnIleAspGluLeuGluGluThr 100
QY 1116 AATAAAGCCTTCAGAAAGGCGAGAGGAGCTCCAGGAGCTGAGAGAGAAAATTGCCAAA 1175
Db 101 AsnArgSerLeuArgLysAlaGluGluGluLeuGlnAspIleLysGluLysIleSerLys 120
QY 1176 GGGAAATGTGGAAGAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTCGCGCAAGCGCGTCTT 1235
Db 121 GlyGluTyrGlyAsnAlaGlyIleMetAlaGluVal----- 132

QY 1236 GAGATGGAGGCAAGGATGAAGAGATCACGAAGACCGAGGCCAGTGCCGGGAGCTGAAG 1295
Db 133 -----GluGluLeuIleLysMetGluGluGlnCysArgAspLeuAsn 146
QY 1296 AAGAAGCTCCAAGAGGAAGAACACACAGCAAGAACTTAGACTAGAAGTGGAGAAGCTG 1355
Db 147 LysArgLeuGluArgGluThrLeuGlnSerLysAspPheLysLeuGluValGluLysLeu 166
QY 1356 CAGAAGAGGATGTCTGAGCTGGAGAAGCTGGAGGAAGCGTTACGCGGAGTAAGTCGGAA 1415
Db 167 SerLysArgIleMetAlaLeuGluLysLeuGluAspAlaPheAsnLysSerLysGlnGlu 186
QY 1416 TGCACCCAGCTCCATCTGAACTGGAGAAGGAGAAGAACCTAAACCAAGACCTGCTGAAC 1475
Db 187 CysTyrSerLeuLysCysAsnLeuGluLysGluArgMetThrThrLysGlnLeuSerGln 206
QY 1476 GAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAG 1535
Db 207 GluLeuGluSerLeuLysValArgIleLysGluLeuGluAlaIleGluSerArgLeuGlu 226
QY 1536 AAGGCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAGTCTTCACTGTGATGCTG 1595
Db 227 LysThrGluPheThrLeuLysGluAspLeuThrLysLeuLysThrLeuThrValMetPhe 246
QY 1596 GTGGATGAGAGGAAAAATATGATGGAGAAAAATAAGCAAGAAGAGAGGAAAGTGGATGG 1655
Db 247 ValAspGluArgLysThrMetSerGluLysThrLysLysLysThrGluAspLysLeuGlnAla 266
QY 1656 TTGAATAAAAACTTTAAGGTGAGCAGGGAAGATCATGGATGTGACGGAAAAAGCTAATC 1715
Db 267 AlaSerSerGlnLeuGlnValGluGlnAsnLysValThrThrValThrGluLysLeuIle 286
QY 1716 GAGGAAAGCAAGAAGCTTTTAAAACTCAAACTCTGAATGGAGGAAAAAGGATACAGTCTG 1775
Db 287 GluGluThrLysArgAlaLeuLysSerLysThrAspValGluGluLysMetTyrSerVal 306
QY 1776 ACAAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGAAGAAGAAAGTCTGTGAA 1835
Db 307 ThrLysGluArgAspLeuLysAsnLysLeuLysAlaGluGluLysGlyAsnAsp 326
QY 1836 CTGAGCTGCAGTGTAGACTTAAGAAAGCGGCTTGATGGCATAGAGGAGGTAGAAAAGG 1895
Db 327 LeuLeuSerArgValAsnMetLeuLysAsnArgLeuGlnSerLeuGluAlaIleGluLys 346
QY 1896 GAAATAAACCGAGGTAGG-----TCGTGCAAGGGGTCTGAGTTACCTGCCCG 1943
Db 347 AspPheLeuLysAsnLysLeuAsnGlnAspSerGlyLysSerThrThrAlaLeuHisGln 366
QY 1944 GAAGACAATAAGATCAGAGAACTAACGCTTGAATCGAGAGACTGAAGAAACGCTCCAG 2003
Db 367 GluAsnAsnLysIleLysGluLeuSerGlnGluValGluArgLeuLysLeuLysLeuLys 386
QY 2004 CAGTTGGAGGTGGGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAG 2063
Db 387 AspMetLysAlaIleGluAspAspLeuMetLysThrGluAspGluTyrGluThrLeuGlu 406
QY 2064 CAGAAGTTCAGAACCGAGGAGGATAAGGCAAACTTCTCTCCAGCAGCTCGAGGAAATC 2123
Db 407 ArgArgTyrAlaAsnGluArgAspLysAlaGlnPheLeuSerLysGluLeuGluHisVal 426
QY 2124 AAACACCAATGGCCCAAGCACAAAGCCATAGAGAAAGGGAGGCCGTGAGCCAGGAAGCC 2183
Db 427 LysMetGluLeuAlaLysTyrLysLeuAlaGluLysThrGlu---ThrSerHisGluGln 445
QY 2184 GAACTGCGACACAGGTTTCGGCTGGAGGAGGCTAAAAGTCGTGATTTACAGGCCGAGGTG 2243
Db 446 TrpLeuPheLysArgLeuGlnGluGluAlaLysSerGlyHisLeuSerArgGluVal 465
QY 2244 CAGGCTCTCAAGGAGAGATCCACGAGCTGATGAACAAGGAAGACCAAGCTGCTCAGCTC 2303
Db 466 AspAlaLeuLysGluLysIleHisGluTyrMetAlaThrGluAspLeuIleCysHisLeu 485
QY 2304 CAAGTCGACTATTCCGGTCTCTCAGCAAAAGATTTATGGAAGAAGAAACTAAGAACAAGAAC 2363

Db 486 GlnGlyAspHisSerValCysLysLysLysLeuAsnGlnGlnGluAsnArgAsnArgAsp 505
QY 2364 ATGGGGAGGGAGGTCTCTCAATCTGACCAAGGAGCTAGAGCTTTCCAAGCGCTACAGCCGA 2423
Db 506 LeuGlyArgGluIleGluAsnLeuThrLysGluLeuGluArgTyrArgHisPheSerLys 525
QY 2424 GCTCTCAGGCCGAGTGGGAAACGGCCGAAGGATGGTGGACGTGCCTGTGGCCTCCACTGGG 2483
Db 526 SerLeuArgProSerLeuAsnGlyArgArgIleSerAspProGlnValPheSerLysGlu 545
QY 2484 GTGCAGACCGAGGGGTGTGCGGGGATGTCTGCGGAGGAGAGACCCCG----- 2531
Db 546 ValGlnThrGluAlaVal-----AspAsnGluProProAspTyrLysSer 560
QY 2532 -----GCTGTGTTTCAATTCGCAAAATCCTTCCAGGAG---GAAATCAC 2570
Db 561 LeuIleProLeuGluArgAlaValIleAsnGlyGlnLeuTyrGluGluSerGluAsnGln 580
QY 2571 ATCATGAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCATGGAACGGTCCCTCGGTCCTC 2630
Db 581 Asp-----GluAspProAsnAspGluGlySerValLeu 591
QY 2631 -----GACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCT 2675
Db 592 SerPheLysCysSerGlnSerThrProCysProValAsn-----ArgLysLeu 607
QY 2676 TGGATTCTTTGGATGAGAAAAAGAGAA-----AACGGTCTTCCACTCCGCAG 2723
Db 608 TrpIleProTrpMetLysSerLysGluGlyHisLeuGlnAsnGly----- 622
QY 2724 GAGAAAGGGCCCCAGGCCAAACACGAGGTGCAGGGCACCCCGGGAGCTGGTCTCTAGCACCA 2783
Db 623 LysMetGlnThrLysProAsnAlaAsnPheValGlnProGlyAspLeuValLeuSerHis 642
QY 2784 AAGCAGGGCCAGCCCTACACATCCGTGTGACACCAGATCATGAGAACAGCACTGCCACC 2843
Db 643 ThrProGlyGlnProLeuHisIleLysValThrProAspHisValGlnAsnThrAlaThr 662
QY 2844 CTGGAGATCACAAAGCCCCACATCTGAA-----GAGTTTTTCTCTAGTACCACCGTCATT 2897
Db 663 LeuGluIleThrSerProThrThrGluSerProHisSerTyrThrSerThrAlaValIle 682
QY 2898 CCTACCTTAGGCAACCAAGAAACCAAGAATAACCATATTATTCATCACCCATGTCATGTCG 2957
Db 683 ProAsnCysGlyThrProLysGlnArgIleThrIleLeuGlnAsnAlaSerIleThrPro 702
QY 2958 CAAAAGCCCCAAA---AGTGCAGATCCTACTCTCGGCCCCAGAACGAGCCATGTCCTGTGTC 3014
Db 703 ValLysSerLysThrSerThrGluAspLeuMetAsnLeuGluGlnGlyMetSerProIle 722
QY 3015 ACGATTACTACTATTTCACAGAGAGAGAGCCCGGAAGGTGGAAGGAGCGCCTTTGCCGAC 3074
Db 723 ThrMetAlaThrPheAlaArgAlaGlnThrProGluSerCysGlySerLeuThrProGlu 742
QY 3075 AGGCCTGCATCC 3086
Db 743 ArgThrMetSer 746
RESULT 13
Q86V48
ID Q86V48 PRELIMINARY; PRT; 1026 AA.
AC Q86V48;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE LÜZPI protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

RC TISSUE=Testis;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L.H., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.F., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC051733; AAH51733.1; -.
DR GO; GO:0005743; C:mitochondrial inner membrane; IEA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001993; Mitoch_carrier.
DR PROSITE; PS00215; MITOCH_CARRIER; 1.
SQ SEQUENCE 1026 AA; 114616 MW; A8A11163CB24DF5A CRC64;

Alignment Scores:
Pred. No.: 2.89e-36 Length: 1026
Score: 893.00 Matches: 283
Percent Similarity: 47.97% Conservative: 143
Best Local Similarity: 31.87% Mismatches: 302
Query Match: 11.55% Indels: 160
DB: 4 Gaps: 32

US-10-788-793-1 (1-4364) x Q86V48 (1-1026)

Qy 1041 CAAGAAATCTCAACCCGCAACTTCGACTCAAACTGGTTGGCTTATCGCAAAGGATTGAG 1100
Db 9 GluThrAlaSerSerArgHisLeuArgPheLysLeuGlnSerLeuSerArgArgLeuAsp 28
Qy 1101 GAGCTGGAAGAGACCAATAAAAGCCTTCAGAAGGCAGAGCAAGAGCTCCAGGAGCTGAGA 1160
Db 29 GluLeuGluGluAlaThrLysAsnLeuGlnLysAlaGluAspGluLeuLeuAspLeuGln 48
Qy 1161 GAGAAATGCCAAAGGGGAATGTGGAACCTCCAGTCTCATGGCGGAAGTGAGAGTCTG 1220
Db 49 AspLysValIleGlnAlaGluGlySerAsnSerSerMetLeuAlaGluIleGluValLeu 68
Qy 1221 CGAAGCGCGTGTGAGATGGAGGCAAGGATGAAGAGATCACGAAGACCGAGGCCAG 1280
Db 69 ArgGlnArgValLeuArgIleGluGlyLysAspGluLysAspGluIleLysArgAlaGluAspLeu 88
Qy 1281 TGCCGGGAGCTGAAGAAGNAGCTCCAAGAGGAAGAACACCACAGCAAGAACTTAGACTA 1340
Db 89 CysArgLeuMetLysGluLysLeuGluGluGluAsnLeuThrArgGluLeuLysSer 108
Qy 1341 GAAGTGGAAGAGCTGCAGAAGAGGATGTCTGAGCTGGAGAAGCTGGAGGAAGCGTTTCAGC 1400
Db 109 GluIleGluArgLeuGlnLysArgMetAlaGluLeuGluLysLeuGluAlaPheSer 128
Qy 1401 CGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTCGAGAAGGAGAGAACCTAACCC 1460
Db 129 ArgSerLysAsnAspCysThrGlnLeuCysLeuSerLeuAsnGluGluArgAsnLeuThr 148

Qy 1461 AAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCC 1520
Db 149 LysLysIleSerSerGluLeuGluMetLeuArgValLysValLysGluLeuGluSerSer 168
Qy 1521 GAGAGTAGACTGGAGAAGCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAGTCC 1580
Db 169 GluAspArgLeuAspLysThrGluGlnSerLeuAlaSerGluLeuGluLysLeuLysSer 188
Qy 1581 TTCACCTGTGATGCTGGTGGATGAGAGGAAAAATATATGATGGAGAAAAATAAGCAAGAAG 1640
Db 189 LeuThrLeuSerPheValSerGluArgLysTyrLeuAsnGluLysGluLysGluAsnGlu 208
Qy 1641 AGAAAGTGGATGGGTGAATAAAACTTTAAGGTGGAGCAGGGAAAAAGTCATCGATGTG 1700
Db 209 LysLeuIleLysGluLeuThrGln-----LysLeuGluGlnAsnLysLysMet----- 224
Qy 1701 ACGGAAAAAGCTAATCGAGGAAAGCAAGAAGCTTTTAAAAACTCAAATCTGAAATGGAGGAA 1760
Db 225 -----AsnArgAspTyrThrArgAsnAlaSerAsnLeuGluArg 237
Qy 1761 AAGGAGTACAGTCTGACAAAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGAAGAA 1820
Db 238 AsnAspLeuArgIle-----GluAspGlyIleSerSerThrLeuProSerLysGlu 254
Qy 1821 GAAAGTCTGTGAACTGAGCTGCAGTGTAGACTTACTAAAGAAGCCGCTTGATGGCATA 1880
Db 255 SerArgArg-----LysGlyGlyLeuAspTyrLeu 264
Qy 1881 GAGGAGGTAGAAAGGGAATA---AACCAGAGTAGTGTGTCGTCGAAGGGCTCTGAGTTCACC 1937
Db 265 LysGlnValGluAsnGluThrArgAsnLysSerGluAsnGluLysAsnArgAsn----- 282
Qy 1938 TGCCCGGAAGACAATAAGATCAGAGAACTAACGCCTTGAATACTGAGAGACTGAAGAAACGG 1997
Db 283 ---GlnGluAspAsnLysValLysAspLeuAsnGlnGluIleGluLysLeuLysThrGln 301
Qy 1998 CTCCAGCAGTTGGAGTGGTGGAGGGGACTTGTATGAAGACCCGAGGACGAATATGACCAG 2057
Db 302 IleLysHisPheGluSerLeuGluGluLeuLysLysMetLysSerLysAsnAsp 321
Qy 2058 TTGGAGCAGAAGTTTCAGAACCCGAGCAGGATAAGGCCAAACTTCTCTCCAGCAGCTCGAG 2117
Db 322 LeuGlnAspAsnTyrLeuSerGluGlnAsnLysAsnLysLeuLeuAlaSerGlnLeuGlu 341
Qy 2118 GAAATCAAAACACCAATGGCCCAAGCACAAAGCCATAGAGAAAGGGGAGGCCGTGAGCCAG 2177
Db 342 GluIleLysLeuGlnIleLysLysGlnLysGluLeuGluAsnGlyGluValGluGlyGlu 361
Qy 2178 GAAGCCGAACTGCGACACAGGTTTCGGCTGGAGGAGGCTAAAAGTCGTGATTTACAGGCC 2237
Db 362 AspAlaPheLeuSerSerLysGlyArgHisGluArgThrLysPheArgGlyHisGlySer 381
Qy 2238 GAGTGCAGGCTCTCAAGGAGAAAGATCCACGAGCTG-----ATGAACAAGGAAGACCAG 2291
Db 382 GluAlaSerValSerLysHisThrAlaArgGluLeuSerProGlnHisLysArgGluArg 401
Qy 2292 CTGTCTCAGCTCCAAGTCGACTATTTCGGTCTTCAGCAAAGATTTTATGGAAGAGAAACT 2351
Db 402 Leu-----ArgAsnArgGluPheAlaLeu 409
Qy 2352 AAGAACAAGAACATGGGGAGGAGGTCTCAATCTGACCAAGGAGCTAGAGCTTTTCCAAG 2411
Db 410 AsnAsnGluAsn-----TyrSerLeuSerAsn 418
Qy 2412 CGGTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAAGGATGGTGGACGTGCCTGTG 2471
Db 419 Arg-----GlnValSerSerProSerPheThrAsnArgArgAla-----AlaLysAla 434
Qy 2472 GCCTCCACTGGGGTCGAGACCGAGCGGGTGTGCGGGGATGTCTCGGAGGAGGAGACCCCG 2531
Db 435 SerHisMetGlyValSerThr-----AspSerGlyThrGlnGluThr--- 448
Qy 2532 GCTGTGTTCAATTCGCAAAATCCTTCCAGGAGGAGGAAAAATCATCATGAGTAATCTTCGACAG 2591

Db 122 ArgSerLysAsnAspCysThrGlnLeuCysLeuSerLeuAsnGluGluArgAsnLeuThr 141
QY 1461 AAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCC 1520
Db 142 LysLysIleSerSerGluLeuGluMetLeuArgValLysValLysGluLeuGluSerSer 161
QY 1521 GAGAGTAGACTGGAGAAAGCCGAGTTAAGCCTCAAAGATCACTTACAAGCTGAAGTCC 1580
Db 162 GluAspArgLeuAspLysThrGluGlnSerLeuAlaSerGluLeuGluLysLeuLysSer 181
QY 1581 TTCACCTGTGATGCTGGTGGATGAGAGGAAAAATATGATGGAGAAAAATAAGCAAGAGAG 1640
Db 182 LeuThrLeuSerPheValSerGluArgLysTyrLeuAsnGluLysGluLysGluAsnGlu 201
QY 1641 AGGAAAGTGGATGGTTGAATAAAAACTTTAAGGTGGAGCGAGGAAAAAGCTCATGATGTG 1700
Db 202 LysLeuIleLysGluLeuThrGln-----LysLeuGluGlnAsnLysLysMet----- 217
QY 1701 ACGGAAAGCTAATCGAGGAAAGCAAGAAAGCTTTTAAAACTCAAATCTGAAATGGAGGAA 1760
Db 218 -----AsnArgAspTyrThrArgAsnAlaSerAsnLeuGluArg 230
QY 1761 AAGGAGTACAGTCTGACAAAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGAAGAA 1820
Db 231 AsnAspLeuAurgile-----GluAspGlyIleSerSerThrLeuProSerLysGlu 247
QY 1821 GAAAGTCTCTGTGAACCTGAGCTGCAGTGTAGACTTACTAAAGAAAGCGGCTTGATGGCATA 1880
Db 248 SerArgArg-----LysGlyGlyLeuAspTyrLeu 257
QY 1881 GAGGAGGTAGAAAGGGAATA---AACCGAGGTAGTTCGTGCAAGGGTCTGAGTTCAACC 1937
Db 258 LysGlnValGluAsnGluThrArgAsnLysSerGluAsnGluLysAsnArgAsn----- 275
QY 1938 TGCCCCGGAAGACAAATAAGATCAGAGAACTAAACGCTTGAATCGAGAGACTGAAGAAACGG 1997
Db 276 ---GlnGluAspAsnLysValLysAspLeuAsnGlnGluIleGluLysLeuLysThrGln 294
QY 1998 CTCCAGCAGTTGGAGGTGGTGGAGGGGACTTGATGARAGACCGAGGACGAATATGACCAG 2057
Db 295 IleLysHisPheGluSerLeuGluGluGluLeuLysLysMetLysSerLysAsnAsnAsp 314
QY 2058 TTGGAGCAGAAAGTTCAGAAACCGAGCAGGATAAGGCAAACTTCCTCTCCAGCAGCTCGAG 2117
Db 315 LeuGlnAspAsnTyrLeuSerGluGluGlnAsnLysAsnLysLeuLeuAlaSerGlnLeuGlu 334
QY 2118 GAAATCAAAACCAAATGGCCCAAGCACAAAGCCATAGAGAAAAGGGAGCGCCGTGAGCCAG 2177
Db 335 GluIleLysLeuGlnIleLysLysGlnLysGluLeuGluAsnGlyGluValGluGlyGlu 354
QY 2178 GAAGCCGAACCTCGCACACAGGTTTCGGCTGGAGGAGGCTAAAGTCGTGATTTACAGGCC 2237
Db 355 AspAlaPheLeuSerSerLysGlyArgHisGluArgThrLysPheArgGlyHisGlySer 374
QY 2238 GAGGTGCAGGCTCTCAAGGAGAAGATCCACGAGCTG-----ATGAACAAGGAAGACCAG 2291
Db 375 GluAlaSerValSerLysHisThrAlaArgGluLeuSerProGlnHisLysArgGluArg 394
QY 2292 CTGTCTCAGCTCCAAGTCGACTATTTCGTCTCTTCAGCAAGAGATTTATGGAAGAAGAACT 2351
Db 395 Leu-----ArgAsnArgGluPheAlaLeu 402
QY 2352 AAGAACAAGAACATGGGGAGGGAGGTCTCAATCTGACCAAGGAGCTAGAGCTTTCCCAAG 2411
Db 403 AsnAsnGluAsn-----TyrSerLeuSerAsn 411
QY 2412 CGTACAGCCGAGCTCTCAGGCCGAGTGGGAAACGGCCGAAGGATGTGGACGTGCTGTG 2471
Db 412 Arg-----GlnValSerSerProSerPheThrAsnArgAla-----AlaLysAla 427
QY 2472 GCCTCCACTGGGGTGCACACCGGAGGGGTGTGCGGGATGCTGCGGAGGAGGAGACCCCG 2531
Db 428 SerHisMetGlyValSerThr-----AspSerGlyThrGlnGluThr--- 441

QY 2532 GCTGTGTTTATTTCGAAATCCTTCCAGGAGGAGAAAAATCACATCATGAGTAATTTTCGACAG 2591
Db 442 -----LysLysThr-----GluAspArgPheValProSerSerLysSer 455
QY 2592 GTAGGCTGAAGAAACCCCATGGAACGGTCTCGTCCGTCGACAGGTTATCCCCCGCAGCG 2651
Db 456 GluGly---LysLysSerArgGluGlnProSerValLeuSerArgTyrProProAlaAla 474
QY 2652 AATGAGCTCACCATGAGGAAGTCTTGGATTCTCTTGGATGAGAAAAAGAGAAACCGTCTCT 2711
Db 475 GlnGluHisSer-----LysAlaTrp-----LysGlyThrSerLysPro 487
QY 2712 TCCACT----- 2717
Db 488 GlyThrGluSerGlyLeuLysGlyLysValGluLysThrThrArgThrPheSerAspThr 507
QY 2718 -----CCGACAGGAGAAAGGGCCCGCCAGGCAAAACCCAGGGTGCAGGGCACCCC 2762
Db 508 ThrHisGlySerValProSerAspProLeuGlyArgAlaAspLysAlaSerAspThrSer 527
QY 2763 GGGAGCTGGTCTTAGCACCAAGCAGGGCCAGCCCTA-----CACATCCGTGTGACA 2816
Db 528 SerGluThrValPheGly---LysArgGlyHisValLeuGlyAsnGlySerGlnValThr 546
QY 2817 CCAGATCATGAGAAACAGCACTGCCACCTGGAGATCACAAGCCCCACATCTGAAGAGTTT 2876
Db 547 ---GlnAlaAlaAsnSerGlyCysSerLysAlaIleGlyAlaLeuAlaSerSerArgArg 565
QY 2877 TTCTCTAGTACCACCGCTATTCTCTACTCTAGGCAACCCAGAAACCAAGAAATAACCATATT 2936
Db 566 SerSerSer-----GluGlyLeuSerLysGlyLysLysAlaAlaAsnGlyLeu 581
QY 2937 CCATCACCCAAT-----GTCTGTGCGAAAAAGCCCCAAAAGT 2972
Db 582 GluAlaAspAsnSerCysProAsnSerLysAlaProValLeuSerLysTyrProTyrSer 601
QY 2973 GCAGAT-----CCTACTCTCGGCCCGCAGAACGAGCCCATG 3005
Db 602 CysArgSerGlnGluAsnIleLeuGlnGlyPheSerThrSerHisLysGluGlyValAsn 621
QY 3006 TCCCTGTTCACGATTACTACTATTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3056
Db 622 GlnProAlaAlaVal-----ValMetGluAspSerSerProHisGluAlaLeuArgCys 639
QY 3057 -----AGGAGCGCTTTGCCGACAGGCTGCATCCCTCCATCCAAATC----- 3098
Db 640 ArgValIleLysSerSerGlyArgGluLysProAspSerAspAspLeuAspIleAla 659
QY 3099 ATGACGCTGTCAACATCTGCAGCTCCCACTGAAATCGTGTCTCTCTCTGAAATCTCAGGAA 3158
Db 660 SerLeuValThrAlaLysLeuValAsnThrThrIleThrProGluProGluProLysPro 679
QY 3159 GTGCTATGGGAGGACTATCTCTCAAAGTCACCCCGGAAAAACAAACT----- 3206
Db 680 GlnProAsnSerArgGluLysAlaLysThrArgGlyAlaProArgThrSerLeuPheGlu 699
QY 3207 -----GTTCCAGCCCCCGTGGGAGAGTACAACTCCAATGCTAAT 3245
Db 700 AsnAspLysAspAlaGlyMetGluAsnGluSerValLysSerValArgAlaSerThrAsn 719
QY 3246 ATCATCACCGAGAGACAAATAAAATTCACATTACCTGGGTTCTCAGTTTAAG---CGA 3302
Db 720 ThrMetGluLeuProAspThrAsn-----GlyAlaGlyValLysSerGln 734
QY 3303 TCTCTGGGCTGCGCTGAAGGCTGAGCCAGTTATCACCGTCCGCTGTCAACGTG 3362
Db 735 ArgProPheSerProArgGluAlaLeuArgSerArgAlaIleIleLysProValIleVal 754
QY 3363 ACAGCGGAGAGGAGGTT-----TCTACAGGCACAGTCTTCGCTCTCCC 3407
Db 755 -----AspLysAspValLysLysIleMetGlyGlySerGlyThrGluThrThrLeuGlu 772

QY	3408	AGGAACCACCTCTCTTCAAGACCCGGTGTAGCAAAAGTGACCAAGCAGCACTATAACTATAACC	3467
Db	773	LysGlnLysProValSerLysProGlyProAsnLysValThrSerSerIleThrIleTyr	792
QY	3468	CCGGTCACAACGTCATCCACACGA	3491
Db	793	ProSerAspSerSerProArg	800
RESULT 15			
Q99NG3			
ID	Q99NG3	PRELIMINARY;	PRT; 1067 AA.
AC	Q99NG3;		
DT	01-JUN-2001	(TrEMBLrel. 17, Created)	
DT	01-JUN-2001	(TrEMBLrel. 17, Last sequence update)	
DT	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)	
DE	Leucine zipper protein 1.		
GN	LuzPl.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ICR;		
RX	MEDLINE=96411647; PubMed=8812416;		
RA	Sun D.S., Chang A.C., Jenkins N.A., Gilbert D.J., Copeland N.G.,		
RA	Chang N.C.;		
RT	"Identification, molecular characterization, and chromosomal		
RT	localization of the cDNA encoding a novel leucine zipper motif-		
RT	containing protein.";		
RL	Genomics 36:54-62(1996).		
DR	EMBL; L49344; AAA98795.1; ..		
DR	HSSP; P04268; 1IC2.		
DR	MGD; MGI:107629; Luzpl.		
SQ	SEQUENCE 1067 AA; 119289 MW; 381F1C16181749CA CRC64;		
Alignment Scores:			
Pred. No.:	1.34e-33	Length:	1067
Score:	839.00	Matches:	275
Percent Similarity:	45.26%	Conservative:	150
Best Local Similarity:	29.29%	Mismatches:	326
Query Match:	10.85%	Indels:	188
DB:	11	Gaps:	28
US-10-788-793-1 (1-4364) x Q99NG3- (1-1067)			
QY	1026	GCCAAATTGGCGAATCAAGAA-----TCTCAAAACCGGCAACTTCGACTCAAACTGGTT	1079
Db	2	AlaGluLeuThrAsnTyrLysAspAlaAlaSerAsnArgHisLeuArgPheLysLeuGln	21
QY	1080	GGCTTATCGCAAGGATTGAGGAGCTGGAAGACCAATAAAAGCCCTTCAGAAGGCAGAG	1139
Db	22	SerLeuSerArgArgLeuAspGluLeuGluGluAlaThrLysAsnLeuGlnArgAlaGlu	41
QY	1140	GAAGAGCTCCAGGAGCTGAGAGAGAAATTCGCAAGGGGAATGTGGAACTCCAGTCTC	1199
Db	42	AspGluLeuLeuAspLeuGlnAspLysValIleGlnAlaGluGlySerAspSerThr	61
QY	1200	ATGGCGGAAGTGAGAGTCTGCGCAAGCGCTGCTTGAGATGGAGGGCAAGGATGAAGAG	1259
Db	62	LeuAlaGluIleGluValLeuArgGlnArgValLeuLysIleGluGlyLysAspGluGlu	81
QY	1260	ATCACGAAGACCGAGCCAGTGCCTGGAGCTGAAGAAGAGCTCCAAGAGGAAGACAC	1319
Db	82	IleLysArgAlaGluAspLeuCysHisThrMetLysGluLysLeuGluGluGluAsn	101
QY	1320	CACAGCAAGGAACCTAGACTAGAAGTGGAGAACTGCAGAAGAGGATGTCTGAGCTGGAG	1379
Db	102	LeuThrArgGluLeuLysSerGluIleGluArgLeuGlnLysArgMetValAspLeuGlu	121
QY	1380	AAGCTGGAGGAACGTTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTG	1439
Db	122	LysLeuGluGluAlaLeuSerArgSerLysAsnGluCysSerGlnLeuLeuCysLeuSerLeu	141

QY	1440	GAGAAGGAGAGAACCTAACCAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGA	1499
Db	142	AsnGluGluArgAsnLeuThrLysLysIleSerSerGluLeuGluMetLeuArgValLys	161
QY	1500	GTTAAAGAACTCAATGCTCCGAGAGTAGACTGGAGAAAGCCGAGTTAAGCCCTCAAAGAT	1559
Db	162	ValLysGluLeuGluSerSerGluAspArgLeuAspLysThrGluGlnSerLeuValSer	181
QY	1560	GACCTTACAAAGCTGAAGTCCTTCACTGTGATGCTGGTGGATGAGAGGAAAAATATGATG	1619
Db	182	GluLeuGluLysLeuLysSerLeuThrLeuSerPheValAsnGluArgLysTyrLeuAsn	201
QY	1620	GAGAAAAATAAGCAAGAGAGAGGAAAGTGGATGGGTTGAATAAAAACTTTAAGGTGGAG	1679
Db	202	GluLysGluLysGluAsnGluLys-----	209
QY	1680	CAGGAAAAAGTCATGGATGTGACGGAAAAAGCTAATCGAGGAAAGCAAGAGCTTTAAAAA	1739
Db	210	-----IleIleLysGluLeuThrGlnLysLeu--GluGlnAsnLysLysMet-----	224
QY	1740	CTCAAACTCTGAAATGGAGGAAAAAGGAGTACAGTCTGACAAAGGAGAGGGATGAGCTGATG	1799
Db	225	-----AsnArgAspHisMetArgAsnAlaSerThrPheLeuGluArgAsnAsp-----	240
QY	1800	GGTAAACTGAGGAGCGAAGAAAGAAAGGTCTCTGTGAACCTGAGCTGCAGTGTAGACTTACTA	1859
Db	241	-----LeuArgIleGluAspGlyIleSerSerThrLeuSerSerLysGluSerLysArg	258
QY	1860	AAGAAGCGGTTGATGGCATAGAGGAGGTAGAAAGGGAATAAACCCGAGGTAGGTTCGTGC	1919
Db	259	LysGlySerLeuAspTyrLeuLysGlnValGluAsnGlu--ThrArgAspLysSer---	276
QY	1920	AAGGGTCTGAGTTCACCTGCCCCGGAAGACAATAAGATCAGAGAACTAACGCTTGAATC	1979
Db	277	GluAsnGluLysAsnArgAsnGlnGluAspAsnLysValLysAspLeuAsnGlnGluIle	296
QY	1980	GAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTGGTGGAGGGGACTTTGATGAAGACC	2039
Db	297	GluLysLeuLysThrGlnIleLysHisPheGluSerLeuGluGluLeuLysLysMet	316
QY	2040	GAGGACGAATATGACCAGTTGGAGCAGAAAGTTTCAGAACCCGAGCAGGATAAGGCAAACTTC	2099
Db	317	ArgAlaLysAsnAsnAspLeuGlnAspAsnTyrLeuThrGluLeuAsnArgAsnArgSer	336
QY	2100	CTCTCCCAGCAGCTCGAGGAAATCAACACCAAAATGGCCAAGCACAAAGCCATAGAGAAA	2159
Db	337	LeuAlaSerGlnLeuGluGluIleLysLeuGlnValArgLysGlnLysGluLeuGlyAsn	356
QY	2160	GGGAGGCCGTGAGCCAGGAAGCCGAACCTGCGACACAGGTTTCGGCTGGAGGAGGCTAAA	2219
Db	357	GlyAspIleGluGlyGluAspAlaPheLeuLeuGlyArgGlyArgHisGluArgThrLys	376
QY	2220	AGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAGGAGAAGATCCACGAGCTG-----	2273
Db	377	LeuLysGlyHisGlySerGluAlaSerValSerLysHisThrSerArgGluLeuSerPro	396
QY	2274	ATGAACAAGGAAGACCAGCTGTCTCTCAGCTCCAAGTCGACTATTCCGTCCTTCAGCAAAGA	2333
Db	397	GlnHisLysArgGluArgLeu-----	404
QY	2334	TTTATGGAAGAAAGAACTAAGAACAAACATGGGAGGGAGGTCTCAATCTGACCAAG	2393
Db	405	AsnArgGluPheAlaLeuSerAsnGluHisTyrSerLeuSerSerLysGlnAlaSerSer	424
QY	2394	GAGCTAGAGCTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAAGG	2453
Db	425	ProValPheThrAsnLysArgAlaAlaLysAla-----	435
QY	2454	ATGGTGGACGTGCCTGTGGCCCTCCACTGGGGTGCAGACCCGAGCGGTTCGCGGGGATGCT	2513
Db	436	-----SerAsnMetGlyMetGlyThrAspSer-----GlyThrGln	447

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 7, 2004, 18:03:38 ; Search time 2349 Seconds
(without alignments)
7892.353 Million cell updates/sec

Title: US-10-788-793-1
Perfect score: 4364
Sequence: 1 ccactgggtcttctaaggga.....aaaaaaaaaaaaaaaaaaaa 4364

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Geneseqn1980a:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4364	100.0	4364	7	ACC45354 Rat L-FIL
2	3666.4	84.0	3785	7	ACC45355 Rat S-FIL
3	3025.6	69.3	4892	8	ADB49155 Novel hum
4	3024	69.3	4892	4	AAI59172 Human pol
5	2987.2	68.5	4247	7	ACC45356 Human L-F
6	2431.2	55.7	3558	4	AAH14619 Human cDN
7	1729.4	39.6	2594	4	AAH14449 Human cDN
8	1527.6	35.0	2402	8	ACH03975 Human cDN
9	790.2	18.1	3415	8	ACC83929 GPBP-inte
10	788.6	18.1	3416	8	ACC83930 GPBP-inte
11	788.6	18.1	3430	8	ACC83928 GPBP-inte
12	777.6	17.8	3998	8	ACC83927 GPBP-inte
13	630.8	14.5	3053	7	ACC90628 Human CGD
14	581.6	13.3	844	4	AAH06432 Human cDN
15	570.2	13.1	1326	4	AAI60958 Human pol
16	560	12.8	808	4	AAH06919 Human cDN
17	559.2	12.8	3587	7	ACA04011 cDNA down
18	541.6	12.4	3025	6	ABK35549 Gene DOC1
19	541.6	12.4	3025	6	ABK64739 Human ben
20	541.6	12.4	3025	7	ABZ71968 Human ova
21	541.6	12.4	3025	7	ACF12851 Human cer
22	496.6	11.4	2355	8	ACC83934 GPBP-inte
23	365.4	8.4	1241	4	AAF22838 Human pro

24	292	6.7	1050	8	ACC83926	Acc83926 GPBP-inte
25	265.4	6.1	795	8	ACC83925	Acc83925 GPBP-inte
26	248.2	5.7	1568	3	AAF18033	Aaf18033 Lung canc
27	240.6	5.5	3537	9	ADC30104	Adc30104 Human nov
28	240.6	5.5	3715	7	ABQ77428	Abq77428 Human CGD
29	240.6	5.5	4848	7	ACF34447	Acf34447 Gene enco
30	239.8	5.5	1625	9	ADB63229	Adb63229 Human CDN
31	231.4	5.3	715	4	AAF22892	Aaf22892 Human pro
32	212.6	4.9	725	4	AAF22947	Aaf22947 Human pro
33	204.4	4.7	720	8	ACC83940	Acc83940 GPBP-inte
34	202.8	4.6	720	8	ACC83924	Acc83924 GPBP-inte
35	195.2	4.5	339	5	AAS33804	Aas33804 Human CDN
36	182.8	4.2	426	4	AAI21417	Aai21417 Probe #11
37	182.8	4.2	426	4	AAI46706	Aai46706 Probe #15
38	182.8	4.2	426	4	ABA48586	Aba48586 Human bre
39	182.8	4.2	426	4	ABA33560	Aba33560 Probe #12
40	182.8	4.2	426	4	AAK40654	Aak40654 Human bon
41	182.8	4.2	426	4	AAK14921	Aak14921 Human bra
42	182.8	4.2	426	4	ABS40222	Abs40222 Human liv
43	182.8	4.2	426	5	AAI07111	Aai07111 Probe #71
C 44	177.2	4.1	547	4	AAH11941	Aah11941 Human CDN
45	154	3.5	1152	8	ACC83941	Acc83941 GPBP-inte

ALIGNMENTS

RESULT 1
ACC45354
ID ACC45354 standard; cDNA; 4364 BP.
XX
AC ACC45354;
XX
DT 18-JUN-2003 (first entry)
XX
DE Rat L-FILIP encoding cDNA SEQ ID NO:1.
XX
KW L-FILIP; S-FILIP; filamin-interacting protein; cell migration;
KW cell death; cytostatic; neuroprotective; immunosuppressive; cancer;
KW tumour metastasis; transplantation therapy; gene; ss.
XX
OS Rattus norvegicus.
XX
FH Key Location/Qualifiers
CDS 75..3713
FT /*tag= a
FT /product= "L-FILIP"
FT /note= "filamin-interacting protein"
XX
PN WO2003018804-A1.
XX
PD 06-MAR-2003.
XX
PF 29-JUL-2002; 2002WO-JP007676.
XX
PR 27-AUG-2001; 2001JP-00256910.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI Sato M, Nagano T;
XX
DR WPI; 2003-268423/26.
DR P-PSDB; ABP97029.
XX
PT Proteins controlling cell migration and cell death and their encoded
PT DNAs, applicable in developing drugs for treating or suppressing cancer
PT or tumor metastasis or as regulators of cell migration for
transplantation.
XX
PS Claim 2; Page 45-54; 96pp; Japanese.
XX
CC The present sequence encodes rat L-FILIP which is a filamin-interacting
CC protein. L-FILIP has a function of controlling cell migration and cell

CC	death. L-FILIP has cytostatic, neuroprotective and immunosuppressive
CC	activities. The L-FILIP protein can be used for controlling cell
CC	migration and cell death, which is applicable in developing drugs for
CC	treating or suppressing cancer or tumour metastasis or as regulators of
CC	cell migration for transplantation therapy, and also for controlling the
CC	mobility and cell death of nerve cells, promoting decomposition of the
CC	actin-binding protein e.g. filamin-interacting protein in the treatment
CC	of preiventrilcular nodular heterotopia
XX	
SQ	Sequence 4364 BP; 1387 A; 988 C; 1165 G; 824 T; 0 U; 0 Other;
	Query Match 100.0%; Score 4364; DB 7; Length 4364;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 4364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 CCACTGGGTCTTCAAGGGATAAAACCAGCGCGAAAGAACACACCAATTGGTTAAGGATC 60
DB	
QY	1 CCACTGGGTCTTCAAGGGATAAACACGCGCGAAAGAACACCAATTGGTTAAGGATC 60
DB	
QY	61 GACAAACAGGTGGGAATGAGATCACGAAATCAAGGTGGAGAAAGTTTCATCTAACCGGCATG 120
DB	
QY	61 GACAAACAGGTGGGAATGAGATCACGAAATCAAGGTGGAGAAAGTTTCATCTAACCGGCATG 120
DB	
QY	121 TCTCCTGCCCCAAGTCTCTCATCATCAGCAGTGATGGTGGTAAGGCCCTCAGAAAGATG 180
DB	
QY	121 TCTCCTGCCCCAAGTCTCTCATCATCAGCAGTGATGGTGGTAAGGCCCTCAGAAAGATG 180
DB	
QY	181 CAAAAAAGAACCAAGGCCAATCGGAAGGAGGAGGATGTCTATGGCTTCCGGAACATCAAAA 240
DB	
QY	181 CAAAAAAGAACCAAGGCCAATCGGAAGGAGGAGGATGTCTATGGCTTCCGGAACATCAAAA 240
DB	
QY	241 GGCACCTCAAAACCATCTGGAGAAAAGTGAGAAAAAGACTAAAGAGTCTGTGGAGTTATCCA 300
DB	
QY	241 GGCACCTCAAAACCATCTGGAGAAAAGTGAGAAAAAGACTAAAGAGTCTGTGGAGTTATCCA 300
DB	
QY	301 AGGAGGACCTCATCCAGTCTCTGAGTATCATGGAAAGGAGATTGCAGGCTCGAGAAATG 360
DB	
QY	301 AGGAGGACCTCATCCAGTCTCTGAGTATCATGGAAAGGAGATTGCAGGCTCGAGAAATG 360
DB	
QY	361 TCATCCACATGCTGAGGACAGAGAAAACCAAGCCCCAGGTTCTGGAGGCACACTATGGAT 420
DB	
QY	361 TCATCCACATGCTGAGGACAGAGAAAACCAAGCCCCAGGTTCTGGAGGCACACTATGGAT 420
DB	
QY	421 CTGCAGAACTTGAGAAAGTGCTTCGGGTCTTGCAACCGAGATGCCATCCTTGCTCAAAGAGA 480
DB	
QY	421 CTGCAGAACTTGAGAAAGTGCTTCGGGTCTTGCAACCGAGATGCCATCCTTGCTCAAAGAGA 480
DB	
QY	481 AGTCCATAGGAGAAGACGCTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAAA 540
DB	
QY	481 AGTCCATAGGAGAAGACGCTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAAA 540
DB	
QY	541 AGCAGAAAGGAGACGTACCGCCGCATGCTAGAGCAGCTGCTGGCTGAGAAAGTGTCAACA 600
DB	
QY	541 AGCAGAAAGGAGACGTACCGCCGCATGCTAGAGCAGCTGCTGGCTGAGAAAGTGTCAACA 600
DB	
QY	601 GCGCACCGTGTAACGAGCTGGAGAACGAGAAAGCAAGCACACTGACTACATGAACAAAGA 660
DB	
QY	601 GCGCACCGTGTAACGAGCTGGAGAACGAGAAAGCAAGCACACTGACTACATGAACAAAGA 660
DB	
QY	661 GCGACGACTTCAACCAACCTGCTGGAGCAGGAGCGAGAGAGGTTGAAAAAGCTCCTTGAAC 720
DB	
QY	661 GCGACGACTTCAACCAACCTGCTGGAGCAGGAGCGAGAGAGGTTGAAAAAGCTCCTTGAAC 720
DB	
QY	721 AAGAAAAGCTTACCAAGCCCGCAAGAAAAGGAAACGCTAAGCGGCTCAACAACTTC 780
DB	
QY	721 AAGAAAAGCTTACCAAGCCCGCAAGAAAAGGAAACGCTAAGCGGCTCAACAACTTC 780
DB	
QY	781 GAGATGAGCTTGTGAGCTCAAGTCTTTCGCCCTCATGTTGGTGGACGAGAGGCAGATGC 840
DB	
QY	781 GAGATGAGCTTGTGAGCTCAAGTCTTTCGCCCTCATGTTGGTGGACGAGAGGCAGATGC 840
DB	
QY	841 ACATCAGCAACTGGGCCTGCAGAGTCAGAAAAGTCCAGGACCTCACTCAGAAAGCTCAGGG 900
DB	

DB	841 ACATCAGCAAACTGGGCCTGCAGAGTCAGAGTCAGAAAAGTCCAGGACCTCACTCAGAAAAGCTGAGGG 900
QY	901 AGGAGGAAAGAAAACCTCAAAGCGGTCACTTACAAATCCAAGGAAGACCGCCAGAAAGCTGC 960
DB	
QY	901 AGGAGGAAAGAAAACCTCAAAGCGGTCACTTACAAATCCAAGGAAGACCGCCAGAAAGCTGC 960
DB	
QY	961 TCAAGTTAGAAAGTGGACTTCGAAACACAAGGCCCTCGAGGTTTTCCTCCAGAGCAGCAAGAGAGA 1020
DB	
QY	961 TCAAGTTAGAAAGTGGACTTCGAAACACAAGGCCCTCGAGGTTTTCCTCCAGAGCAGCAAGAGAGA 1020
DB	
QY	1021 TGAACGCCAAAATTGGCGAATCAAGAAATCTCACAAACCGGCAAACTTCGACTCAAACTGGTTG 1080
DB	
QY	1021 TGAACGCCAAAATTGGCGAATCAAGAAATCTCACAAACCGGCAAACTTCGACTCAAACTGGTTG 1080
DB	
QY	1081 GCTTATCGCAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCCTTCAGAAAGGCAGAGG 1140
DB	
QY	1081 GCTTATCGCAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCCTTCAGAAAGGCAGAGG 1140
DB	
QY	1141 AAGAGCTCCAGGAGCTGAGAGAGAGAAAATTGCCAAAGGGGAATGTGGAATCTCCAGTCTCA 1200
DB	
QY	1141 AAGAGCTCCAGGAGCTGAGAGAGAGAAAATTGCCAAAGGGGAATGTGGAATCTCCAGTCTCA 1200
DB	
QY	1201 TGGCGAAAGTGGAGAGTCTGCGCAAGCGCGTGTGAGATGGAGGCAAGGATGAAGAGA 1260
DB	
QY	1201 TGGCGAAAGTGGAGAGTCTGCGCAAGCGCGTGTGAGATGGAGGCAAGGATGAAGAGA 1260
DB	
QY	1261 TCACGAAGACCCGAGGCCAGTGCCTCCGGAGCTGAAGAAGAAGCTCCAAGAGGAAGAACACC 1320
DB	
QY	1261 TCACGAAGACCCGAGGCCAGTGCCTCCGGAGCTGAAGAAGAAGCTCCAAGAGGAAGAACACC 1320
DB	
QY	1321 ACAGCAAGGAACTTAGACTAGAAAGTGAGAGAAAGCTGCAGAAAGAGGATGTCTGAGCTGGAGA 1380
DB	
QY	1321 ACAGCAAGGAACTTAGACTAGAAAGTGAGAGAAAGCTGCAGAAAGAGGATGTCTGAGCTGGAGA 1380
DB	
QY	1381 AGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGG 1440
DB	
QY	1381 AGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGG 1440
DB	
QY	1441 AGAAGGAGAAAGAACCTAAACCAAGACCTGCTGAACGAGCTGGAGTGGTCAAGAGTCGAG 1500
DB	
QY	1441 AGAAGGAGAAAGAACCTAAACCAAGACCTGCTGAACGAGCTGGAGTGGTCAAGAGTCGAG 1500
DB	
QY	1501 TTAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAAGCCGAGTTAAGCCTCAAAGATG 1560
DB	
QY	1501 TTAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAAGCCGAGTTAAGCCTCAAAGATG 1560
DB	
QY	1561 ACCTTACAAAAGCTGAAAGTCCTTCACTGTGATGCTGGTGGATGAGAGGAAAAATATGATGG 1620
DB	
QY	1561 ACCTTACAAAAGCTGAAAGTCCTTCACTGTGATGCTGGTGGATGAGAGGAAAAATATGATGG 1620
DB	
QY	1621 AGAAAATAAGCAAGAGAGAGGAAAGTGGATGGGTTGAATAAAAACCTTTAAGGTGAGC 1680
DB	
QY	1621 AGAAAATAAGCAAGAGAGAGGAAAGTGGATGGGTTGAATAAAAACCTTTAAGGTGAGC 1680
DB	
QY	1681 AGGAAAAAGTCAATGGATGTGACGGAAGAGCTAATCGAGGAAAGCAAGAAAGCTTTTAAAAAC 1740
DB	
QY	1681 AGGAAAAAGTCAATGGATGTGACGGAAGAGCTAATCGAGGAAAGCAAGAAAGCTTTTAAAAAC 1740
DB	
QY	1741 TCAAACTGAAATGGAGGAAAGAGGATACAGTCTGACAAAAGGAGGGGATGAGCTGATGG 1800
DB	
QY	1741 TCAAACTGAAATGGAGGAAAGAGGATACAGTCTGACAAAAGGAGGGGATGAGCTGATGG 1800
DB	
QY	1801 GTAAACTGAGGACGGAAGAAAGGTCCTGTGAACCTGAGCTGAGTGTAGACTTACTAA 1860
DB	
QY	1801 GTAAACTGAGGACGGAAGAAAGGTCCTGTGAACCTGAGCTGAGTGTAGACTTACTAA 1860
DB	
QY	1861 AGAAGCGGCTTATGGCATAGAGGAGGTAGAAAAGGGGAAATAAACCGAGGTAGGTCGTGCA 1920
DB	
QY	1861 AGAAGCGGCTTATGGCATAGAGGAGGTAGAAAAGGGGAAATAAACCGAGGTAGGTCGTGCA 1920
DB	
QY	1921 AGGGGTCTGAGTTCACCTGCCCGGAAGACATAAGATCAGAGAACTAACCGCTTGAAATCG 1980
DB	
QY	1921 AGGGGTCTGAGTTCACCTGCCCGGAAGACATAAGATCAGAGAACTAACCGCTTGAAATCG 1980
DB	

Db ||||| 4141 TTTTGAATAGGACAGAGTTTAACAGTTGTGCATTTTGCACTATCAAGCCATGAGTTTGAT 4200
Qy ||||| 4201 ATATGGGTTATAAGAAAGAAATACTTTTCAGAGCTATCACAGGGTCTCTAAACTTTTGGAA 4260
Db ||||| 4201 ATATGGGTTATAAGAAAGAAATACTTTTCAGAGCTATCACAGGGTCTCTAAACTTTTGGAA 4260
Qy ||||| 4261 AAACAAAGCCCTTAATATGACCTCAGGNAACAATTTGNACATGAATAAAATGGAATG 4320
Db ||||| 4261 AAACAAAGCCCTTAATATGACCTCAGGAAACAATTTGAACATGAATAAAATGGAATG 4320
Qy ||||| 4321 AACTGTGGAATCTTTAAAAAANAAAAAANAAAAAANAAAAA 4364
Db ||||| 4321 AACTGTGGAATCTTTAAAAAANAAAAAANAAAAAANAAAAA 4364

RESULT 2
ACC45355
ID ACC45355 standard; cDNA; 3785 BP.
XX
AC ACC45355;
XX
DT 18-JUN-2003 (first entry)
XX
DE Rat S-FILIP encoding cDNA SEQ ID NO:3.
XX
XW L-FILIP; S-FILIP; filamin-interacting protein; cell migration;
KW cell death; cytostatic; neuroprotective; immunosuppressive; cancer;
KW tumour metastasis; transplantation therapy; gene; ss.
OS Rattus norvegicus.

XX Key Location/Qualifiers
FH CDS 237..3134
FT /*tag= a
FT /product= "S-FILIP"
FT /note= "filamin-interacting protein"
XX
PN WO2003018804-A1.

XX
PD 06-MAR-2003.
XX
XX 29-JUL-2002; 2002WO-JP007676.
PF
XX 27-AUG-2001; 2001JP-00256910.
PR
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
PA
XX Sato M, Nagano T;
PI WPI; 2003-268423/26.
XX P-PSDB; ABP97030.
DR
XX Proteins controlling cell migration and cell death and their encoded
PT DNAs, applicable in developing drugs for treating or suppressing cancer
PT or tumor metastasis or as regulators of cell migration for
PT transplantation.
XX Claim 5; Page 60-68; 96pp; Japanese.

XX The present sequence encodes rat S-FILIP which is a filamin-interacting
CC protein. S-FILIP has a function of controlling cell migration and cell
CC death. S-FILIP has cytostatic, neuroprotective and immunosuppressive
CC activities. The S-FILIP protein can be used for controlling cell
CC migration and cell death, which is applicable in developing drugs for
CC treating or suppressing cancer or tumour metastasis or as regulators of
CC cell migration for transplantation therapy, and also for controlling the
CC mobility and cell death of nerve cells, promoting decomposition of the
CC actin-binding protein e.g. filamin-interacting protein in the treatment
XX of preiventricular nodular heterotopia

Sequence 3785 BP; 1207 A; 853 C; 990 G; 735 T; 0 U; 0 Other;

Query Match 84.0%; Score 3666.4; DB 7; Length 3785;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3667; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 697 AGAGGTTGAAAAAGCTCCTTTGAACAAGAAAAAGCTTACCAAGCCCGAAAAAGGAAA 756
Db 118 AAAGGTTGAAAAAGCTCCTTTGAACAAGAAAAAGCTTACCAAGCCCGAAAAAGGAAA 177
Qy 757 ACGCTAAGCGGCTCAACAAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCA 816
Db 178 ACGCTAAGCGGCTCAACAAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCA 237
Qy 817 TGTTCGTGGACGAGAGGCGAGATGCACATCGAGCAACTGGGCCCTGCAGAGTCAGAAAAGTCC 876
Db 238 TGTTCGTGGACGAGAGGCGAGATGCACATCGAGCAACTGGGCCCTGCAGAGTCAGAAAAGTCC 297
Qy 877 AGGACCTCACTCAGAACTGAGGGAGGAGGAAGAAAACTCAAAGCGGTCACTTACAAAT 936
Db 298 AGGACCTCACTCAGAACTGAGGGAGGAGGAAGAAAACTCAAAGCGGTCACTTACAAAT 357
Qy 937 CCAAGGAAGACCCGCCAGAAAGCTGCTCAAAGTTAGAAGTGACTTCGAACACAAAGGCCTCGA 996
Db 358 CCAAGGAAGACCCGCCAGAAAGCTGCTCAAAGTTAGAAGTGACTTCGAACACAAAGGCCTCGA 417
Qy 997 GGTTCCTCCAGGACGACGAAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCACAAAC 1056
Db 418 GGTTCCTCCAGGACGACGAAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCACAAAC 477
Qy 1057 GCGAACTTCGACTCAAACTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGAAGAGACCA 1116
Db 478 GCGAACTTCGACTCAAACTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGAAGAGACCA 537
Qy 1117 ATAAAGCCCTTCAGAAAGCGCAGAGGAAGAGCTCCAGAGCTGAGAGAGAAAAATTGCCAAAG 1176
Db 538 ATAAAGCCCTTCAGAAAGCGCAGAGGAAGAGCTCCAGAGCTGAGAGAGAAAAATTGCCAAAG 597
Qy 1177 GGGATGTGGAAACTCCAGTCTCATGGCGGAAAGTGAGAGAGTCTGCGCAAGCGCGTCTTG 1236
Db 598 GGGATGTGGAAACTCCAGTCTCATGGCGGAAAGTGAGAGAGTCTGCGCAAGCGCGTCTTG 657
Qy 1237 AGATGGAGGGCAAGGATGAAGATCACGAAGACCGAGGCCCCAGTCGCGGAGCTGAAGA 1296
Db 658 AGATGGAGGGCAAGGATGAAGATCACGAAGACCGAGGCCCCAGTCGCGGAGCTGAAGA 717
Qy 1297 AGAAGCTCCAAGAGGAAGAACACCCAGCAAGGAACCTTAGACTAGAAAGTGGAAGCTGC 1356
Db 718 AGAAGCTCCAAGAGGAAGAACACCCAGCAAGGAACCTTAGACTAGAAAGTGGAAGCTGC 777
Qy 1357 AGAAGAGGATGCTCGAGCTGGAGAAAGCTGGAGGAAGCGTTAGCCGGAGTAAAGTCGGAAT 1416
Db 778 AGAAGAGGATGCTCGAGCTGGAGAAAGCTGGAGGAAGCGTTAGCCGGAGTAAAGTCGGAAT 837
Qy 1417 GCACCCAGCTCCATCTGAACCTGGAGAGGAGAAAGAACCTTAACCAAGACCTGCTGAACG 1476
Db 838 GCACCCAGCTCCATCTGAACCTGGAGAGGAGAAAGAACCTTAACCAAGACCTGCTGAACG 897
Qy 1477 AGCTGAGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGA 1536
Db 898 AGCTGAGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGA 957
Qy 1537 AGCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAGTCCTTCACTGTGATGCTGG 1596
Db 958 AGCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAGTCCTTCACTGTGATGCTGG 1017
Qy 1597 TGGATGAGAGGAAAAATATGATGGAGAAAAATAAGCAAGAAAGAGAGGAAAAAGTGGATGGGT 1656
Db 1018 TGGATGAGAGGAAAAATATGATGGAGAAAAATAAGCAAGAAAGAGAGGAAAAAGTGGATGGGT 1077
Qy 1657 TGAATAAAAACTTTAAGGTGGAGCGGGAAGAAAGTATGGATGTGACGGAAAAAGCTAATCG 1716
Db 1078 TGAATAAAAACTTTAAGGTGGAGCGGGAAGAAAGTATGGATGTGACGGAAAAAGCTAATCG 1137
Qy 1717 AGGAAAGCAAGAAGCTTTTAAAACTCAAATCTGAAATGGAGGAAAAAGGAGTACAGTCTGA 1776

Db 1138 AGGAAAGCAAGAAGCTTTTAAAACTCAAATCTGAAATGAGGAAAAAGGAGTACAGTCTGA 1197
QY 1777 CAAAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGAAGAAGAAAGGTCCTGTGAAC 1836
Db 1198 CAAAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGAAGAAGAAAGGTCCTGTGAAC 1257
QY 1837 TGAGCTGCAGTGTAGACTTTACTAAAGAACGGCTTGATGGCATAGAGGAGGTAGAAAGGG 1896
Db 1258 TGAGCTGCAGTGTAGACTTTACTAAAGAACGGCTTGATGGCATAGAGGAGGTAGAAAGGG 1317
QY 1897 AAATAAACCGAGGTAGTTCGTGCAAGGGGTCTGAGTTTACCTGCCGGAAGACAAATAAGA 1956
Db 1318 AAATAAACCGAGGTAGTTCGTGCAAGGGGTCTGAGTTTACCTGCCGGAAGACAAATAAGA 1377
QY 1957 TCAGAGAACTAACGCTTGAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAAGGTGG 2016
Db 1378 TCAGAGAACTAACGCTTGAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAAGGTGG 1437
QY 2017 TGGAGGGGACTTGTATGAAGACCGGAGGACGAATATGACAGTTGGAGCAGAAGTTCAGAA 2076
Db 1438 TGGAGGGGACTTGTATGAAGACCGGAGGACGAATATGACAGTTGGAGCAGAAGTTCAGAA 1497
QY 2077 CCGAGCAGGATAAGGCAAACTTCTCTCCCAGCAGCTCGAGGAAATCAAACACCAAATGG 2136
Db 1498 CCGAGCAGGATAAGGCAAACTTCTCTCCCAGCAGCTCGAGGAAATCAAACACCAAATGG 1557
QY 2137 CCAAGCACAAAGCCATAGAGAAAGGGGAGGCCGTGAGCCAGGAAGCCGAACCTGCGACACA 2196
Db 1558 CCAAGCACAAAGCCATAGAGAAAGGGGAGGCCGTGAGCCAGGAAGCCGAACCTGCGACACA 1617
QY 2197 GGTTCGGCTGGAGGAGGCTAAAAAGTCGTGATTTACAGGCCGAGGTGACGGCTCTCAAGG 2256
Db 1618 GGTTCGGCTGGAGGAGGCTAAAAAGTCGTGATTTACAGGCCGAGGTGACGGCTCTCAAGG 1677
QY 2257 AGAAGATCCACGAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTATT 2316
Db 1678 AGAAGATCCACGAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTATT 1737
QY 2317 CGGTCTTCAGCAAAAGATTATGGAAGAAGAACTAAAGAACAGAACATGGGGAGGGAGG 2376
Db 1738 CGGTCTTCAGCAAAAGATTATGGAAGAAGAACTAAAGAACAGAACATGGGGAGGGAGG 1797
QY 2377 TCCTCAATCTGACCAAGGAGCTAGAGCTTTTCCAAAGCGCTACAGCCGAGCTCTCAGGCCGA 2436
Db 1798 TCCTCAATCTGACCAAGGAGCTAGAGCTTTTCCAAAGCGCTACAGCCGAGCTCTCAGGCCGA 1857
QY 2437 GTGGGAA CGGCCGAAAGGATGGTGGACGTGCCCTGTGGCTCCACTGGGTGACAGCCGAGG 2496
Db 1858 GTGGGAA CGGCCGAAAGGATGGTGGACGTGCCCTGTGGCTCCACTGGGTGACAGCCGAGG 1917
QY 2497 CGGTGTGCGGGGATGCTGCGGAGGAGGAGACCCCGGCTGTGTTTCAATTCGCAATCCTTCC 2556
Db 1918 CGGTGTGCGGGGATGCTGCGGAGGAGGAGACCCCGGCTGTGTTTCAATTCGCAATCCTTCC 1977
QY 2557 AGGAGGAAAAATCACATCATGAGTAATCTTCGACAGGTAGGCCCTGAAGAAACCCATGGAAC 2616
Db 1978 AGGAGGAAAAATCACATCATGAGTAATCTTCGACAGGTAGGCCCTGAAGAAACCCATGGAAC 2037
QY 2617 GGTCTCGGTCTCGACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTT 2676
Db 2038 GGTCTCGGTCTCGACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTT 2097
QY 2677 GGATTCCTTGGATGAGAAAAAGAGAAAAACGGTCTTCCACTCCGACGAGGAAGGGCCCA 2736
Db 2098 GGATTCCTTGGATGAGAAAAAGAGAAAAACGGTCTTCCACTCCGACGAGGAAGGGCCCA 2157
QY 2737 GGCCAAACCGAGGTGACGGGCACCCCGGGAGCTGGTCTTAGCACCAAGCAGGGCCAGC 2796
Db 2158 GGCCAAACCGAGGTGACGGGCACCCCGGGAGCTGGTCTTAGCACCAAGCAGGGCCAGC 2217
QY 2797 CCCTACACATCCGTGTGACACCAGATCATGAGAACAGCACTGCCACCCCTGGAGATCACAA 2856

Db 2218 CCCTACACATCCGTGTGACACCAAGATCATGAGAAACAGCACTGCCACCCCTGGAGATCACAA 2277
QY 2857 GCCCCACATCTGAAGAGTTTTTCTCTAGTACCACCGTCATTTCTACCTTAGGCAACAGA 2916
Db 2278 GCCCCACATCTGAAGAGTTTTTCTCTAGTACCACCGTCATTTCTACCTTAGGCAACAGA 2337
QY 2917 AACCAAGAATAACCATTTATTCATCACCCCAATGTGCATGTGCAAAAAGCCCCAAAAGTGCAG 2976
Db 2338 AACCAAGAATAACCATTTATTCATCACCCCAATGTGCATGTGCAAAAAGCCCCAAAAGTGCAG 2397
QY 2977 ATCCTACTCTCGGCCCCAGAACGAGCCCATGTCCCCTGTACAGATTACTACTATTTCAGAG 3036
Db 2398 ATCCTACTCTCGGCCCCAGAACGAGCCCATGTCCCCTGTACGATTACTACTATTTCAGAG 2457
QY 3037 AGAAGAGCCCCGGAAGGTGGAAGGAGCGCCTTTGCGCAGAGGCCCTGCATCCCCCATCCAAA 3096
Db 2458 AGAAGAGCCCCGGAAGGTGGAAGGAGCGCCTTTGCGCAGAGGCCCTGCATCCCCCATCCAAA 2517
QY 3097 TCATGACGGTGTCAACATCTGCAGTCTCCACTGCAAAATCGCTGTCTCTCTGAAATCTCAGG 3156
Db 2518 TCATGACGGTGTCAACATCTGCAGTCTCCACTGCAAAATCGCTGTCTCTCTGAAATCTCAGG 2577
QY 3157 AAGTGCCTATGGGAAGGACTATCTCAAAGTCACCCCGAAAAACAAAACCTGTTCAGGCC 3216
Db 2578 AAGTGCCTATGGGAAGGACTATCTCAAAGTCACCCCGAAAAACAAAACCTGTTCAGGCC 2637
QY 3217 CCGTGGGAAGTACAACTCCAATGCTAATATCATCACCGAAGACAAATAAAATTACAA 3276
Db 2638 CCGTGGGAAGTACAACTCCAATGCTAATATCATCACCGAAGACAAATAAAATTACAA 2697
QY 3277 TTCACCTGGGTTCTCAGTTTAAAGCGATCTCTTGGGCCCTGCCGTGAAGGCGTGAGCCCCAG 3336
Db 2698 TTCACCTGGGTTCTCAGTTTAAAGCGATCTCTTGGGCCCTGCCGTGAAGGCGTGAGCCCCAG 2757
QY 3337 TTATCACCGTCCGGCTGTCAA CGTGACAGCGGAGAAAGAGGTTTCTACAGGCACAGTCC 3396
Db 2758 TTATCACCGTCCGGCTGTCAA CGTGACAGCGGAGAAAGAGGTTTCTACAGGCACAGTCC 2817
QY 3397 TTCGCTCTCCAGGAACCACTCTCTTCAAGACCCCGGTGCTAGCAAAAGTGACAGCACTA 3456
Db 2818 TTCGCTCTCCAGGAACCACTCTCTTCAAGACCCCGGTGCTAGCAAAAGTGACAGCACTA 2877
QY 3457 TAACTATAACCCCGGTCAAA CGTCATCCACAGAGGAACCCCAATCAGTGTCAAGGACAAG 3516
Db 2878 TAACTATAACCCCGGTCAAA CGTCATCCACAGAGGAACCCCAATCAGTGTCAAGGACAAG 2937
QY 3517 ATGGGTCACTCTCAGCGCCCTACCCCCACCCCGCATTCCTATGTCAAAAAGGTATGAAAGCTG 3576
Db 2938 ATGGGTCACTCTCAGCGCCCTACCCCCACCCCGCATTCCTATGTCAAAAAGGTATGAAAGCTG 2997
QY 3577 GAAAGCCAGTAGTGGCAGCCTCAGGAGCAGGAAATCTGACCAAAATTCAGCCTCGAGCTG 3636
Db 2998 GAAAGCCAGTAGTGGCAGCCTCAGGAGCAGGAAATCTGACCAAAATTCAGCCTCGAGCTG 3057
QY 3637 AGACTCAGTCTATGAAAAATAGAGCTGAAGAAATCTGCAGCCAGCACTGCCTCTCTTG 3696
Db 3058 AGACTCAGTCTATGAAAAATAGAGCTGAAGAAATCTGCAGCCAGCACTGCCTCTCTTG 3117
QY 3697 GAGGGGGAAGGCTGAGGCAGTGGCTAAGGGGTATGTTGTAAGGATGCTACTGCTGC 3756
Db 3118 GAGGGGGAAGGCTGAGGCAGTGGCTAAGGGGTATGTTGTAAGGATGCTACTGCTGC 3177
QY 3757 AGTGGAAACAAAACCTTCTCTGTGCCAACCCCTTCTTGACTACTAAATTTAAGTTTAA 3816
Db 3178 AGTGGAAACAAAACCTTCTCTGTGCCAACCCCTTCTTGACTACTAAATTTAAGTTTAA 3237
QY 3817 ATATCTTGTTTATAAAATAAACCATTTAATAGCCATGCACCCCTCCCATTTTGTGCATC 3876
Db 3238 ATATCTTGTTTATAAAATAAACCATTTAATAGCCATGCACCCCTCCCATTTTGTGCATC 3297
QY 3877 TGTTCATGAGGGGAATAGAAATTAATTAGCAGAAATTTCTGTTTGTGTAATGTTCTGTT 3936
Db 3298 TGTTCATGAGGGGAATAGAAATTAATTAGCAGAAATTTCTGTTTGTGTAATGTTCTGTT 3357

Db 1398 AATGCACATTGAACAACTTGGCCTGCAAAGCCAGAAAGTACAGGATCTTACTCAGAAGCT 1457

QY 896 GAGGGAGGAGGAAGAAAACCTCAAAGCGGTCACTTACAAATCCAAGGAAGACCGCCAGAA 955

Db 1458 GAGGGAAGAGAGAGAGCTCAAAGCCATTACTTCCAATCCAAGAGACAGACAGAA 1517

QY 956 GCTGCTCAAGTTAGAGTGGACTTCGAACACAAGGCCTCGAGGTTTTTCCAGGAGCAGCA 1015

Db 1518 ATTGCTCAAGTTAGAGTGGACTTTGAACACACAGGCTTCGAGGTTTTCTCAAGAGCATGA 1577

QY 1016 AGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCACACCCGGCAAACCTTCGACTCAAACT 1075

Db 1578 AGAGATGAACGCTAAACTGGCTAATCAAGAGTCTCACAAATAGGCAACTTAGACTCAAGCT 1637

QY 1076 GGTGGCTTATCGCAAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAGGC 1135

Db 1638 GGTGGCTTAACCCCAAAGAATCGAGGAGCTAGAAGAGACCAACAAAATCTGCAGAAGGC 1697

QY 1136 AGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAAACTCCAG 1195

Db 1698 AGAGGAAGAACTTCAAGAAATTAAGAGATAAAATTTGCCAAAGGAGAATGTGGAAACTCTAG 1757

QY 1196 TCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTGTGATGAGATGGAGGGCAAGGATGA 1255

Db 1758 CCTCATGGCAGAAGTGGAAAAATCTTCGAAAGCGTGTGCTTGAATTGAAGGTAAAGATGA 1817

QY 1256 AGAGATCACGAAGACCGAGGCCAGTGCAGGAGCTGAAGAAGAAAGCTCCAAGAGGAAGA 1315

Db 1818 GGAGATCACTAAAACTGAATCCAGTGTAGGGAATTTAGGGAAGTGTGCAAGAGGAAGA 1877

QY 1316 ACACCACAGCAAGGAACCTTAGACTAGAAAGTGGAGAGCTGCAGAAAGAGGATGTCTGAGCT 1375

Db 1878 ACACCATAGTAAGGAGCTCAGACTTGAAGTTGAGAACTACAGAAGAGAATGTCTGAACT 1937

QY 1376 GGAGAAGCTGGAGGAAGCGTTCAGCCGGAGTAAGTCGAATGCACCCAGCTCCATCTGAA 1435

Db 1938 AGAGAAATTTGGAAGAAGCATTTAGCAAGAGTAATCTGAGTGCACCCAGCTACATTTAA 1997

QY 1436 CCTGGAGAAGGAAGAACCTTAACCAAAGACCTGTGAAACGAGCTGGAGGTGGTCAAGAG 1495

Db 1998 TCTGGAGAAAGAAAGAACTTAACCAAAGACCTGTCTAAATGAATTTGGAGGTGGTCAAGAG 2057

QY 1496 TCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAGGCCGAGTTAAGCCTCAA 1555

Db 2058 TCGAGTTAAAGAAATTTGGAATGTTCTGAAAGTAGATTGGAAGAGGCTGAATTAAGCCTAAA 2117

QY 1556 AGATGACCTTACAAAGCTGAAGTCTTCACTGTGATGCTGGTGGATGAGAGGAAAAATAT 1615

Db 2118 AGATGATCTTACCAAGTTGAAGTCATTTACCGTGATGCTGGTTGATGAAAGGAAAAATAT 2177

QY 1616 GATGGAGAAAAATAAGCAAGAGAGAGGAAAGTGGATGGTTGAATAAAAACTTTAAGGT 1675

Db 2178 GATGGAAAAATAAAACAAGAGAGAGAAAAAGTGGATGGACTCAATAAAAAATTTAAGGT 2237

QY 1676 GGAGCAGGGAAAAAGTCATGGATGTGACGGAAAAAGCTAATCGAGGAAAGCAAGAGCTTTT 1735

Db 2238 GGAACAAGGAAAAAGTTATGGATGTAACTGAAAAAACTAATTTGAAGAAAGTAAGAAACTTTT 2297

QY 1736 AAAACTCAAATCTGAAATGGAGGAAAGGAGTACAGTCTGACAAAGGAGAGGATGAGCT 1795

Db 2298 AAAACTAAAATCTGAAATGGAGGAAAAAGTATACAACTTGACAAGAGAAAGAGATGAGTT 2357

QY 1796 GATGGGTAAACTGAGGAGCGGAAGAAAGGTCCTGTGAACTGAGCTGCAGTGTAGACTT 1855

Db 2358 GATAGGCAAATTTGAAAAGTGAAGAAGAAAAATCCTCTGAATTAAGCTGCAGTGTGACTT 2417

QY 1856 ACTAAGAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGGGAATAAACCAGGAGTAGGTC 1915

Db 2418 ACTAAGAAGAGACITTGATGGTATAGAGGAAGTGGAAAGAGAAATAACAAGAGGAAGGTC 2477

QY 1916 GTGCAAGGGGTCTGAGTTACCTGCCCGGAAGACAAATAAGATCAGAGAACTAACGCTTGA 1975

Db 2478 ACGAAAAGGGTCTGAGCTCACCTGCCCGGAAGATAATAAGATTAAAGAACTTAACACTTGA 2537

QY 1976 AATCGAGAGACTGAAGAAAACGGCTCCAGCAGTTTGGAGGTGGAGGGGACTTGTATGAA 2035

Db 2538 AATTGAGAGACTGAAGAAAACGTCTCCAACAATTTGGAAGTGGTGAAGGGATTTGTATGAA 2597

QY 2036 GACCGAGGACGAATATGACCAGTTGGAGCAGAAAGTTTCAGAAACCGAGCAGGATAAGGCCAAA 2095

Db 2598 GACAGAAGATGAGTATGATCAGCTGGAACAGAAATTTAGAACTGAGCAGGATAAGGCTAA 2657

QY 2096 CTTCTCTCCCAGCAGCTCGAGGAAATCAAAACACCAAAATGGCCAAAGCACAAAAGCCATAGA 2155

Db 2658 CTTCTCTCTCAACAACATAGAGGAGATCAAGCACCAAAATTTGCCAAGAAATAAAGCAATAGA 2717

QY 2156 GAAAGGGGAGGCCGTGAGCCAGGAAGCCGAACTTGCGACACAGGTTTTCGGCTGGAGGAGGC 2215

Db 2718 GAAAGGTGAGGTTGTGAGCCAGGAAGCTGAACCTGAGACACAGATTTTCGGTTGGAAGAAGC 2777

QY 2216 TAAAGTCTGTATTTACAGGCCGAGGTGCAGGCTCTCAAGGAGAGATCCACGAGCTGAT 2275

Db 2778 TAAAGTCTGAGACTTAAAAAGCCGAAGTACAAGCTCTTAAAGAGAAGATTCACGAATTAAT 2837

QY 2276 GAAACAAGGAAGACCACTGTCTCAGCTCCAAGTCCGACTATTCGGTCTCTCAGCAAAAGATT 2335

Db 2838 GAACAAAGAAGATCAGCTTCTCAGCTCCAGGTAGATTAATCTGTACTTCAACAAAGATT 2897

QY 2336 TATGGAAGAAAGAACTAAGAAACAAGAACATGGGAGGGAGGTCTCTCAATCTGACCAAGGA 2395

Db 2898 TATGGAAGAAAGAAATAAGAAACAAAACATGGGCGCAGAGGTTCTCAATCTGACCAAGA 2957

QY 2396 GCTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGAAACGGCCGAGGAT 2455

Db 2958 GTTGAGCTTTCCAAGCGCTACAGCAGAGCTCTTAGGCCCAGTGTGAATGGAAGAAGAAAT 3017

QY 2456 GGTGGACGTGCCTGTGBCCTCCACTGGGGTGCAGACCGAGGCGGTGTGCGGGGATGCTGC 2515

Db 3018 GGTGGATGTTCTCTGACGTCAACTGGAGTCCAAACTGATGCAGTCAGCGTGAAGCAGC 3077

QY 2516 GGAGGAGGAGACCCCGCTGTGTTCAATTCGCAAAATCCTTCCAGGAGGAAAAATCAGATCAT 2575

Db 3078 AGAGGAAGAAACGCCAGCTGTATTATACGGAATCTTCCAGGAAGAAAAATCATATTAT 3137

QY 2576 GAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCCATGGAACGGTCTCTCGGTCTCTGACAG 2635

Db 3138 GAGTAATCTTCGGCAGGTGGGATTGAAGAAACCCGTGGAAAGATCTTCTGTTCTAGACAG 3197

QY 2636 GTATCCCCCAGCAGCAATGAGCTCACCATGAGAAAGTCTTTGGATTCTCTTGGATGAGAAA 2695

Db 3198 GTATCTCTCAGCAGCAAAATGAGCTCACTATGAGAAAAGTCTTTGGATTCCATGGATGAGAAA 3257

QY 2696 AAGAGAAAAACGGTCTTCCACTCCGACGAGAGAAAGGGCCCCAGGCCAAAAACAGGTTGCAGG 2755

Db 3258 GAGGAAAAACGGCCCTCCATCACTCAGGAGAAAGGGCCCCGAAACAAATTCAGTCCAGG 3317

QY 2756 GCACCCCGGGAGCTGGTCTTAGCACCCAAAGCAGGGCCAGCCCCCTACACATCCGTGTGAC 2815

Db 3318 GCACCCAGGAGAGGTAGTCTTTTACCACAAAGCAGGGCCAGCCCCCTGCATATTCGAGTGAC 3377

QY 2816 ACCAGATCATGAGAAACAGCACTGCCACCCCTGGAGATCACAAGCCCCACATCTGAAAGATT 2875

Db 3378 ACCAGACCAACGAGAACAGCACTGGACCTTTGGAGATAACAAGCCCCGACATCTGAAAGAAAT 3437

QY 2876 TTTCTCTAGTACCAACCGTCATTCTTACCTTAGGCAACCCAGAAAACCAAGAAATAACCATTA 2935

Db 3438 TTTTCTAGTACCACTGTCTTCTTCTTCTAGGGAATCAGAAAACCAAGAAATAACCATTA 3497

QY 2936 TCCATCACCCCAATGTCAATGTGCAAAAAAGCCAAAAAGTGCAGATCCTACTCTCGGCCCAGA 2995

Db 3498 TCCATCACCAAAAGCTTATGCCTCAAAAAACAAAAAGTGGAGATACTACTCTTTGGCCCAGA 3557

QY 2996 ACGAGCCATGTCCTCTGTACGATTACTACTATTTTCCAGAGAGAAGAGCCCGGAAGGTGG 3055

Db 3558 ACGAGCCATGTCCCAGTCACAATTTACTACATTTTCCAGAGAGAAGACTCCAGAAAGTGG 3617

CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the									
CC	utilisation of the activities such as: Immune system suppression,									
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic									
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,									
CC	assays for receptor activity, arthritis and inflammation, leukaemias and									
CC	C.N.S disorders. Note: The sequence data for this patent did not form									
CC	part of the printed specification									
XX										
SQ	Sequence 4892 BP; 1689 A; 987 C; 1145 G; 1071 T; 0 U; 0 Other;									
	Query Match	69.3%;	Score 3024;	DB 4;	Length 4892;					
	Best Local Similarity	82.5%;	Pred. No. 0;							
	Matches 3592;	Conservative	0;	Mismatches 735;	Indels 27;	Gaps 10;				
QY	1	CCACTGGGTTCTTCAAGGGATAAAACCGCGGC-GAAAGAACACACCATTGGTT-AAGGAG	58							
Db	558	CCACTGGGTTCTTCAAGGGATAAACTACCTACATAGAGACATACCTTTGGTTAAAGGAG	617							
QY	59	TCGACAAACAGGTGGGAATGAGATCACGAAATCAAGTGGAGAAAGTTTCATCTAACGGGCA	118							
Db	618	CTGCCGTCAAGTGGGAATGAGATCTCGAAACCAAGGTGGTGAAGTGCACTGTATGGGCA	677							
QY	119	TGTCTCCTGCCCAAGTCTCCATCATCAGCAGTGAATGGTGGTAAGGCCCTCAGAAGA	178							
Db	678	TATCTCCTGTCCCAAGCCTCCATCATCGGCAATGCTGGTGAATAAAGTCTCTCAGAAGA	737							
QY	179	TGCAAAAAAGAACAAAG--GCCAATCGGAAGGAGGAGGATGTATGGCTTCGCGAACTAT	235							
Db	738	TGCAAAAAAGAAGAAATCAAAATAGGAAGGAGGATGATGTATGGCCTCAGGAACGT	797							
QY	236	CAAAAGGCACCTCAAAACCATCTGGAGAAAGTGAGAAAAAGACTAAGAAGTCTGTGGAGTT	295							
Db	798	CAAACGACACCTAAAAACATCTGGAGATGTGAACGAAAAAACTAAGAAATCCCTGGAGTT	857							
QY	296	ATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGGAGTTGCAGGCTCGAGA	355							
Db	858	ATCCAAGAAGACCTCATCCAACTACTCAGTATAATGGAAGGGGAGTTGCAAGCCAGAGA	917							
QY	356	AGATGTCATCCACATGCTGAGGACAGAGAAAAACCAAGCCCGAGGTTCTGGAGGCACACTA	415							
Db	918	AGATGTGATCCACATGCTGAAGACAGAGAAAAACCAAGCCTGAGGTTCTGGAGGCTCATTA	977							
QY	416	TGGATCTGCAGAACTGAGAAAGTGCTTCGGGTCTTGCAACCGAGATGCCATCCTTGCTCA	475							
Db	978	CGGGTCTGCGAGCCAGAGAAAGTGCTGCGGGTCTTGCAACCGAGATGCCATTCTTGCCCA	1037							
QY	476	AGAGAAGTCCATAGGAGAAGACGTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGA	535							
Db	1038	GGAGAAATCCATAGGAGAAGATGTCTATGAGAAACCGATTTTCAGAGCTGGACAGACTTGA	1097							
QY	536	GGAAAGCAGAAGGAGACGTACCGCCGCATGCTAGACAGCTGCTGCTGGCTGAGAAGTG	595							
Db	1098	GGAAAAACAGAAAGAAACCTACCGCGCATGCTAGACAGCTGTTGCTGGCCGAGAAGTG	1157							
QY	596	TCACAGGCGCACCGTGACGAGCTGGAGAACGAGAAACGAGAACACACACTGACTACATGAA	655							
Db	1158	TCATAGGCGCACCGTATACGAGTTAGAGAACGAGAAAGCATAAAACACACTGACTACATGAA	1217							
QY	656	CAAGAGCGACGACTTCACCAAACCTGCTGGAGCAGGACGAGAGGTTGAAAAAGCTCCT	715							
Db	1218	CAAGAGCGACGACTTCACCAAACCTGCTGGAGCAGGACGGGAGAGGTTAAAAAAGCTCCT	1277							
QY	716	TGAACAAGAAAAAGCTTACCAAGCCCGCAAGAAAAAGAAAAACGCTAAGCGGCTCAACAA	775							
Db	1278	TGAACAAGAAAAAGGCTTATCAAGCCCGCAAGAAAAAGAAAAATGCTAAACGACTCAATAA	1337							
QY	776	ACTTCGAGATGAGCTTGTGAAGCTCAAGTCCCTTCGCCCTCATGTTGGTGGACGAGAGGCA	835							
Db	1338	ACTAAGAGATGAGCTTGTAAACTCAAAATCCCTTTGCACTCATGCTGGTGGATGAAAGACA	1397							
QY	836	GATGCACATCGAGCAACTGGGCCCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAGCT	895							
Db	1398	AATGCACATTGAACAACCTTGGCCCTGCAAGCCAGAAAGTACAGGATCTTTACTCAGAAGCT	1457							

QY	896	GAGGAGGAGGAAGAAAAAAGCTCAAAGCGGTCACTTACAAAATCCAAGGAAGACCGCCAGAA	955
Db	1458	GAGGAAGAAGAAGAGAAGCTCAAAGCCATTACTTTCCAAAATCCAAGAAGACAGACAGAA	1517
QY	956	GCTGCTCAAGTTTAGAAGTGGACTTCGAACACAAAGSCCTCGAGGTTTTCCCAAGGACGACA	1015
Db	1518	ATTGCTCAAGTTTAGAAGTGGACTTTGAAACACAAGGCTTCGAGGTTTTCTCAAGAGCATGA	1577
QY	1016	AGAGATGAACGCCCAAAATTGGCGAATCAAGAATCTCACAAACCGGCAACTTCGACTCAAACT	1075
Db	1578	AGAGATGAACCGCTAAACTGGCTAATCAAGAGTCTCACAAATAGGCAACTTAGACTCAAGCT	1637
QY	1076	GGTTGGCTTATCGCAAGGATTGAGGAGCTGGAGAGACCATAAAGCCCTTCAGAGGCG	1135
Db	1638	GGTTGGCTTAACCCAAAGAATCGAGGAGCTAGAGAGACCACAACAAAATCTGCAGAGGCG	1697
QY	1136	AGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCAAAAGGGGAATGTGAAAACCTCCAG	1195
Db	1698	AGAGGAAGAACTTCAAGAAATTAAGAGATAAAATTGCCAAAAGGAGAATGTGGAACACTCTAG	1757
QY	1196	TCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTCTTGAGATGGAGGCAAGGATGA	1255
Db	1758	CCTCATGGCAGAAGTGGAAAAATCTTCGAAAGCGGTGCTTGAATGGAAGGTAAAGATGA	1817
QY	1256	AGAGATCACGAAGACCGAGGCCAGTGCCCGGAGCTGAAGAAAGAGCTCCAAGAGGAAGA	1315
Db	1818	GGAGATCACTAAACTGAATCCCAGTGTAGGGAATTGAGGAAGAGCTGCAAGAGGAAGA	1877
QY	1316	ACACCACAGCAAGGAACTTAGACTAGAAGTGGAGAAGCTGCAGAAGAGGATGTCTGAGCT	1375
Db	1878	ACACCATAGTAAGGAGCTCAGACTTGAAGTTGAAAGTTGAAAGCTACAGAAGAGAAATGTCTGAACT	1937
QY	1376	GGAGAAAGCTGGAGGAAGCGTTTACGCCGAGTGAAGTCCGGAATGCACCCAGCTCCATCTGAA	1435
Db	1938	AGAGAAATTTGGAAGAAGCACTTTAGCAAGAGTAAATCTGAGTGCACCCAGCTACATTTAAA	1997
QY	1436	CCTGGAGAAGGAGAAGAAACCTTAACCAAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAG	1495
Db	1998	TCTGGAGAAAGAAAGAACTTAACCAAAGACCTGCTTAAATGAATTGGAGGTGGTCAAGAG	2057
QY	1496	TCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAAGCCGAGTTAAGCCTCAA	1555
Db	2058	TCGAGTTAAAGAAATTGGAATGTTCTGAAAGTAGATTGGAAGAGGCTGAATTAGCCTAAA	2117
QY	1556	AGATGACCTTACAAAGCTGAAAGTCTTCACTGTGATGCTGGTGGATGAGAGGAATAATAT	1615
Db	2118	AGATGATCTTACCAAGTTGAAGTCAATTTACCGTGATGCTGGTTGATGAAAGGAAAAATAT	2177
QY	1616	GATGGAGAAAAATAAGCAAGAAGAGAGGAAAGTGGATGGTTGAAATAAAAACTTTAAGGT	1675
Db	2178	GATGGAAAAATAAAAACAAGAAGAGAGAAAGTGGATGGACTCAATAAAAAATTTTAAGGT	2237
QY	1676	GGAGCAGGGAAGAAAGTCAATGGATGTGACGGAAAAAGCTAATCGAGGAAAGCAAGAGCTTTT	1735
Db	2238	GGAAACAAGGAAAGTTATGGATGTAACCTGAAAAAACTAATTGAAGAAAGTAAGAACTTTT	2297
QY	1736	AAAACCTCAAATCTGAATGGAGGAAAAAGGAGTACAGTCTGACAAAGGAGAGGGATGAGCT	1795
Db	2298	AAAACCTAAATCTGAATGGAGGAAAAAGTATACAACTTGACAAAGAGAAAGATGAGTT	2357
QY	1796	GATGGGTAAACTGAGGAGCGAAGAAAGGTCCTGTGAACTGAGCTGCAGTGTAGACTT	1855
Db	2358	GATAGGCAAAATTGAAAAAGTGAAGNAGAAAAATCTCTGAAATTAAGCTGCAGTGTGACTT	2417
QY	1856	ACTAAAGAAGCGGCTTGATGGCATAGAGGAGGTAGAAAAAGGGAATAAAACCGAGGTAGGTC	1915
Db	2418	ACTAAAGAAGAGACTTGATGGTATAGAGGAAAGTGGAAAAAGAGAAATAACAAGAGGAAGGTC	2477
QY	1916	GTGCAAGGGGCTTGAGTTACCTGCCCGGAAGACAAATAAGATCAGAGAACTAACGCTTGA	1975
Db	2478	ACGAAAAGGGTCTGAGCTCACCTGCCCGGAAGATAATAAGATTAAGGAACCTAACACTTGA	2537

Db 4679 AATCCTATTACAAAATRAACACATACTTTAACTATTGTCATTTGCCTCTTTTCACATCATGA 4738
QY 4194 GTTTGATATATGGGTTATAAGAAA-AGAATACTTTTCAGAGCTATCACAGGGTCTCTAAAC 4252
Db 4739 ATTGCTTTATGTGCTGGAANAACATCACATAGCTATCACAGGGCCTGGACCTCTTAAAA 4798
QY 4253 TTTTGGAAAAACAAGCCCTTAATATGACCTCAGGAAACAATTTGAACATGAATAAAA 4312
Db 4799 TTTTGCAAAAACAAGGTTCTAAGATGATTTTCAGGAAATAATGTGAACATGTAATAAAA 4858
QY 4313 TGGAAATGAACCTGTGGAATCTTAAAAAATAAAAA 4346
Db 4859 TGGAAATGAATATGGAATCTTAAAAAATAAAAA 4892

RESULT 5
ACC45356
ID ACC45356 standard; cDNA; 4247 BP.
XX
AC ACC45356;
XX
DT 18-JUN-2003 (first entry)
XX
DE Human L-FILIP encoding cDNA SEQ ID NO:5.
XX
KW L-FILIP; S-FILIP; filamin-interacting protein; cell migration; cancer;
KW cell death; cytostatic; neuroprotective; immunosuppressive; cancer;
KW tumour metastasis; transplantation therapy; gene; ss.
XX
OS Homo sapiens.
XX

FH Key Location/Qualifiers
FT CDS 7..3648
FT /*tag= a
FT /product= "L-FILIP"
FT /note= "filamin-interacting protein"

XX WO2003018804-A1.

XX 06-MAR-2003.

XX 29-JUL-2002; 2002WO-JP007676.

XX 27-AUG-2001; 2001JP-00256910.

XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.

XX Sato M, Nagano T;

XX WPI; 2003-268423/26.

XX P-PSDB; ABP97031.

XX Proteins controlling cell migration and cell death and their encoded
PT DNAs, applicable in developing drugs for treating or suppressing cancer
PT or tumor metastasis or as regulators of cell migration for
PT transplantation.

XX Claim 8; Page 73-81; 96pp; Japanese.

XX The present sequence encodes human L-FILIP which is a filamin-interacting
CC protein. L-FILIP has a function of controlling cell migration and cell
CC death. L-FILIP has cytosolic, neuroprotective and immunosuppressive
CC activities. The L-FILIP protein can be used for controlling cell
CC migration and cell death, which is applicable in developing drugs for
CC treating or suppressing cancer or tumour metastasis or as regulators of
CC cell migration for transplantation therapy, and also for controlling the
CC mobility and cell death of nerve cells, promoting decomposition of the
CC actin-binding protein e.g. filamin-interacting protein in the treatment
CC of preinfiltricular nodular heterotopia

SQ Sequence 4247 BP; 1502 A; 841 C; 997 G; 907 T; 0 U; 0 Other;

Query Match 68.5%; Score 2987.2; DB 7; Length 4247;

Best Local Similarity 82.6%; Pred. No. 0;
Matches 3523; Conservative 0; Mismatches 718; Indels 25; Gaps 8;
QY 69 GTGGGAATGAGATCACGAAATCAAGGTGGAGAAAGTTTCATCTAACGGGCATGTCTCTGC 128
Db 1 GTGGGAATGAGATCTGAAACCAAGGTGGTGAAGTGCATCTGTATGGGCATATCTCTGT 60
QY 129 CCCAAGTCTCCATCATCAGCAGTGTGTTGAAGGGCCCTCAGAAGATGCAGAAAAAG 188
Db 61 CCCAAGCCCTCCATCATCGGCAATGCTGGTGAAGAAAGTCTCTCAGAAGATGCAGAAAAAG 120
QY 189 AACAAAG--GCCAATCGGAAGGAGGAGATGTATGGCTTCCGGAACTATCAAAAGGCAC 245
Db 121 AAGAAGAAATCAAAATAGGAAGGAGGATGTATGTCATGGCCTCAGGAACCTGTCAAACGACAC 180
QY 246 CTCAAAACCATCTGGAGAAAAGTGAGAAAAAGACTAAGAAAGTCTGTGGAGTTATCCAAGGAG 305
Db 181 CTAAAAACATCTGGAGAAATGTGAACGAAAAAACTAAGAAAAATCCCTGGAGTTATCCAAAGAA 240
QY 306 GACCTCATCCAGCTCCTGAGTATCATGGAAGGGAGTTGCAGGCTCGAGAAGATGTCTATC 365
Db 241 GACCTCATCCAACTACTCAGTATAATGGAAGGGAGTTGCAGGCCAGAGAAGATGTGATC 300
QY 366 CACATGCTGAGGACAGAGAAAAACCAAGCCCGAGGTTCTGGAGGCACACTATGATCTTCA 425
Db 301 CACATGCTGAAGACAGAGAAAAACCAAGCCCTGAGGTTCTGGAGGCTCATACGGGTCTGCG 360
QY 426 GAACCTGAGAAAGTCTTCGGGTCTCTGCACCAGAGATGCCATCTTGTCTCAAGAGAAGTCC 485
Db 361 GAGCCAGAGAAAAGTCTGCGGTCTTGCCACCAGAGATGCCATCTTGGCCAGGAGAAATCC 420
QY 486 ATAGGAGAAGACGTCTATGAGAAAACCTATCTCAGAGCTGGACAGACTGGAGGAAAAAGCAG 545
Db 421 ATAGGAGAAGATGTCTATGAGAAAACCGATTTTCAGAGCTGGACAGACTTGAGGAAAAACAG 480
QY 546 AAGGAGACGTACCGCGCATGCTAGAGCAGCTGCTGCTGGTGAGAAAGTGTACAGAGCGC 605
Db 481 AAAGAAACCTACCGCGCATGCTAGAGCAGCTGCTGCTGGCCGAGAAGTGTCTATAGCGC 540
QY 606 ACCGTGTACGAGCTGGAGAACGAGAACGACACACACACTGACTACATGAACAGAGCGAC 665
Db 541 ACCGTATACGAGTTAGAGAACGAGAACGATTAACACACACTGACTACATGAACAGAGCGAC 600
QY 666 GACTTCACCAACCTCTGGAGCAGGAGCGGAGAGGTTGAAAAAGCTCCTTGAACAAGAA 725
Db 601 GACTTCACCAACCTCTGGAGCAGGAGCGGAGAGGTTAAAAAAGCTCCTTGAACAAGAA 660
QY 726 AAAGCTTACCAAGCCGCAAGAAAAAGGAAACGCTAAGCGGCTCAACAACTTCGAGAT 785
Db 661 AAGGCTTATCAAGCCGCAAGAAAAAGGAAAAATGCTAAACGACTCAATAAACTAAGAGAT 720
QY 786 GAGCTTGTGAAGCTCAAGTCTCTTCGCCCTCATGTTGGTGGACGAGAGGCGAGATGCACATC 845
Db 721 GAGCTTGTAAACTCAAACTCCTTTGCACCTCATGCTGGTGGATGAAAGACAAATGACACATT 780
QY 846 GAGCAACTGGGCTGCAGAGTCAGAAAAGTCCAGGACCTCACTCAGAAGCTGAGGAGGAG 905
Db 781 GAACAACTTGGCTGCAAGGCCAGAAAAGTACAGGATCTTACTCAGAAGCTGAGGAGAA 840
QY 906 GAAGAAAAAATCAAGCGGTCACTTACAAATCAAGGAAGACCGCCAGAAAGCTGCTCAAG 965
Db 841 GAAGAGAAGCTCAAGGCCATTACTTCCAAATCCAAAGAAAGACAGACAGAAATGCTCAAG 900
QY 966 TTAGAAGTGGACTTCGAACACAGGCCTCGAGGTTTCCAGGAGCACGAAGAGATGAAC 1025
Db 901 TTAGAAGTGGACTTTGAACACAGGCTTCGAGGTTTCTCAAGAGCATGAAGAGATGAAC 960
QY 1026 GCCAAATTGGCGAATCAAGAAATCTCACAAACCGCAACTTCGACTCAAACTGGTGGCTTA 1085
Db 961 GCTAAACTGGCTAATCAAGAGTCTCACAATAGSCAACTTAGACTCAAGCTGGTGGCTTA 1020
QY 1086 TCGCAAAGGATTTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAGGCCAGAGGAGAG 1145

Db 1021 ACCCAAGAATCGAGGAGCTAGAAGAGACCAACAAAAATCTGCAGAAAGCCAGAGGAAGAA 1080

Qy 1146 CTCCAGAGAGCTGAGAGAGAAAAATTGCCRAAGGGGAATGTGGAAACTCCAGTCTCATGSCG 1205

Db 1081 CTTCAAGAATTAAAGAGATAAAAAATTGCCAAAGGAGAATGTGGAACCTCTAGCCTCATGSCA 1140

Qy 1206 GAAGTGGAGAGTCTCGCAAGCGCGTGTCTTGAGATGGAGGGCAAGGATGAAGAGATCACG 1265

Db 1141 GAAGTGAAAAATCTTCGAAAGCGTGTGCTTGAAATGGAAGGTAAAGATCAGGAGATCACT 1200

Qy 1266 AAGACCGAGGCCAGTGCCGGGAGCTGAAAGAAGAGCTCCAAGAGGAAGAAACCAACAGC 1325

Db 1201 AAAACTGAATCCCAGTGTAGGGAAATTGAGGAAGAAAGCTGCAAGAGGAAGAAACACCATAGT 1260

Qy 1326 AAGGAACCTTAGACTAGAAGTGGAGAAAGCTGCAGAAAGAGGATGTCTGAGCTGGAGAAGCTG 1385

Db 1261 AAGGAGCTCAGACTTGAAGTTGAGAAGCTACAGAAGAGAGAAATGTCTGAACTAGAGAAATTG 1320

Qy 1386 GAGGAAGCGTTACGCCGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAG 1445

Db 1321 GAAGAAGCATTTAGCAAGAGTAAATCTGAGTGCACCCAGCTACATTTTAAATCTGGAGAAA 1380

Qy 1446 GAGAAGAACCTTAACCAAGACCTGCTGTGAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAA 1505

Db 1381 GAAAAGAACTTAACCAAGACCTGCTAAATGAAATTGGAGGTGGTCAAGAGTCGAGTTAAA 1440

Qy 1506 GAACTCGAATGCTCCGAGAGTAGACTGGAGAAGGCCGAGTTAAGCCTCAAAAGATGACCTT 1565

Db 1441 GAATTGGAATGTTCTGAAAGTAGATTGGAAAAGGCTGAATTAAGCCTAAAAAGATGATCTT 1500

Qy 1566 ACAAAGTGAAAGTCTTCACTGTGATGCTGTGGATGAGAGGAAAAATATGATGGAGAAA 1625

Db 1501 ACCAAGTGAAGTCATTTACCGTGATGCTGTGTTGATGAAGGAAAAATATGATGGAAAA 1560

Qy 1626 ATAAAGCAAGAAGAGAGGAAAAGTGGATCGGTTGAATAAAAACTTTTAAGGTGGAGCAGGGA 1685

Db 1561 ATAAAACAAGAAGAGAGAAAAAGTGATGGACTCAATAAAAAATTTTAAGGTGGAACAGGA 1620

Qy 1686 AAAGTCATGGATGTACCGAAAAAGCTAATCGAGGAAAGCAAGAGCTTTTAAAACTCAAA 1745

Db 1621 AAAGTTATGGATGTAACCTGAAAAACTAATTGAAGAAAGTAAGAAACTTTTAAAACTAAA 1680

Qy 1746 TCTGAATGGAGGAAAGGAGTACAGTCTGCACAAAAGGAGGGGATGAGCTGATGGGTAAA 1805

Db 1681 TCTGAATGGAGGAAAAAAGTATACAACCTTGACAAGAGAAAGAGATGAGTTGATAGGCAAA 1740

Qy 1806 CTGAGGACGGAAGAAAGAGTCCCTGTGAACCTGAGCTGCAGTGTAGACTTACTAAAGAAG 1865

Db 1741 TTGAAAAGTGAAGAGAAAAATCCCTCTGAATTAAGCTGCAGTGTGACTTACTAAAGAAG 1800

Qy 1866 CGGCTTGATGGCATAGAGGAGGTAGAAAGGGAAATAAACCGAGGTAGGTGTCGAAGGG 1925

Db 1801 AGACTTGATGGTATAGAGGAAGTGGAAAGAGAAATAACAAGAGGAAGGTACAGAAAAGGG 1860

Qy 1926 TCTGAGTTACCTGCCCGAAGACAATAAGATCAGAGAACTAACCGCTTGAAATCGAGAGA 1985

Db 1861 TCTGAGCTCACCTGCCCGGAAGAATAATAAGATTAAAGGACTAACACTTGAAATTGAGAGA 1920

Qy 1986 CTGAAGAAACGGCTCCAGCAGTTGGAGGTGGTGGAGGGGACTTGATGAAGACCGAGGAC 2045

Db 1921 CTGAAGAAAACGTCTCCAACAATTGGAAGTGGTCGAAGGGGATTTTGATGAAGACAGAGAT 1980

Qy 2046 GAATATGACCCAGTTGGAGCAGAAATTAAGATCAGAAACCGAGCAGGATAAGGCCAAACTTCCTCTCC 2105

Db 1981 GAGTATGATCAGCTGGAACAGAAATTTAGAACTGAGCAGGATAAGGCTAACTTCCTCTCT 2040

Qy 2106 CAGCAGCTCGAGGAAATCAAAACACCAATGGCCCAAGCACAAAAGCCATAGAGAAAAGGGAG 2165

Db 2041 CAACAACTAGAGGAGATCAAGCACCAAAATGCCAAGATAAAAGCAATAGAGAAAGGTGAG 2100

Qy 2166 GCCGTAGCCAGGAAGCCGAACCTGCGACACAGGTTTCGGCTGGAGGAGGCTAAAAATCGT 2225

Db 2101 GTTGTGAGCCAGGAAGCTGAACCTGAGACACAGATTTTCGGTTGGAAGAGCTAAAAAGTCGA 2160

Qy 2226 GATTTACAGGCCGAGGTGCAGGCTCTCAAGGAGAGAAGATCCACGAGCTGATGAACAAGGAA 2285

Db 2161 GACTTAAAAAGCCGAAGTACAAGCTCTTAAAGAGAGAAGATTCAGGAATTAATGAACAAGAA 2220

Qy 2286 GACCAGCTGTCTCAGTCCAAAGTCGACTATTCCGTCTTTCAGCAAAAGATTTATGGAAGAA 2345

Db 2221 GATCAGCTTTCTCAGTCCAGGTAGATTATTCTGTACTTCAACAAAGATTTATGGAAGAA 2280

Qy 2346 GAAACTAAAGAACAAAGAACATGGGAGGGAGGTCTCTCAATCTGACCAAGAGGCTAGAGCTT 2405

Db 2281 GAAAATAAGAACAAAAACATGGGCGAGGAGGTTCTCAATCTGACCAAAAGAGTTGGAGCTT 2340

Qy 2406 TCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGAAACGGCCGAAGGATGGTGGACGTG 2465

Db 2341 TCCAAGCGCTACAGCAGAGCTCTTAGGCCCAGTGTGAATGGAAGAAATGGTGGATGTT 2400

Qy 2466 CCTGTGGCCTCCACTGGGGTGCAGACCGAGGCCGCTGTGCGGGGATGCTGCGGAGGAGGAG 2525

Db 2401 CCTGTGACGTCAACTGGAGTCCAAACTGATGCAGTCAGCGGTGAAGCAGCAGAGGAAGAA 2460

Qy 2526 ACCCGGCTGTGTTCAATTTCGAAATCCTTCCAGGAGGAAAAATCACATCATGAGTAATCTT 2585

Db 2461 ACGCCAGCTGTATTTCATACGGAATCCTTCCAGGAAGAAAAATCATATTATGAGTAATCTT 2520

Qy 2586 CGACAGGTAGGCCCTGAAGAAAACCATGGAAACGGTCTCGGTCTCTCGACAGGTATCCCCCA 2645

Db 2521 CGGAGGTGGGATTGAAGAAAACCCGTGGAAAGATCTTCTGTCTTAGACAGGTATCCTCCA 2580

Qy 2646 GCAGCGAATGAGCTCACCATGAGGAAGTCTTGGATTCTTGGATGAGAAAAAGAGAAAAAC 2705

Db 2581 GCAGAAAATGAGTCTACATATGAGAAAGTCTTGGATTCCATGATGAGAAAAAGGGAAAAAC 2640

Qy 2706 GGTCTTTCACCTCCGAGGAGAAAAGGGCCCAGGCCCAAAACCAAGGTGCAGGGCACCCCGGG 2765

Db 2641 GGCCCTCCATCATCTCAGGAGAAAGGGCCCCGGAACAATAATCCAGTCCAGGGCACCCAGGA 2700

Qy 2766 GAGCTGGTCTTAGCACCAAAGCAGGGCCAGCCCCCTACACATCCCGTGTGACACCCAGATCAT 2825

Db 2701 GAGGTAGTCTTTTACCAAAGCAGGGCCAGCCCCCTGCATATTTCGAGTGACACCAGACCAC 2760

Qy 2826 GAGAACAGCACTGCCACCTGGAGATCACAAGCCCCACACATCTGAAGAGTTTTTCTCTAGT 2885

Db 2761 GAGAACAGCACTGCGACTTTGGAGATAACAAGCCCGACATCTGAAGAAATTTTCTCTAGT 2820

Qy 2886 ACCACGTCATTCCTACCTTAGGCAACCAGAAACCAAGAATAACCATATTTCATCACC 2945

Db 2821 ACCACTGTCAATTCCTTACCTTAGGGRATCAGAAACCAAGAATAACCATATTTCATCACC 2880

Qy 2946 AATGTCATGTGCAAAAAACCCCAAAAGTGCAGATCCTTACTCTCGGCCAGAAACGAGCCATG 3005

Db 2881 AACGTTATGCCCTCAAAAAACAAAAAGTGGAGATACTACTCTTGGCCAGAAACGAGCCATG 2940

Qy 3006 TCCCCTGTACGATTACTACTATTTCCAGAGAGAGAGAGCCCGGAAGGTGGAAGGAGCGCC 3065

Db 2941 TCCCAGTCACAAATTACTACATTTTCCAGAGAGAGAGACTCCAGAAAAGTGGAAAGGCGCA 3000

Qy 3066 TTTGCCGACAGGCCCTGCATCCCCCATCCAAATCATGACGGTGTCAAACATCTGCAGCTCCC 3125

Db 3001 TTTGCAGACAGGCCCAATCCCCCTATTTCAGATAATGACGGTGTCTACATCAGCAGCACCA 3060

Qy 3126 ACTGAAATCGCTGTCTCTCTGAATCTCAGGAAGTGCCTATGGGAAGGACTATCCTCAAA 3185

Db 3061 GCTGAGATTGCAGTTCTCCCGAATCCCAGGAAATGCCCATGGGACGGAACAATCCTCAAA 3120

Qy 3186 GTCACCCCGGAAAAACAAACTGTTCCAGCCCCCGTGGGAAAGTACAACTCCAATGCTAAT 3245

Db 3121 GTCACCCCGAAAAAACAGACTGTTCCAACTCCAGTACGGAATACAACTCCAATGCCAAT 3180

Qy 3246 ATCATCACCAACGGAAGACAATAAAATTCACATTCACCTGGGTTCTCAGTTTAAAGCGATCT 3305

Db 3181 ATCATAAACCAAGAGGACAATAAAATTCACATTCACCTTAGGTTCTCAGTTTAAACGGTCC 3240

QY	813	CTCATGTTGGTGGACGAGAGGCGCAGATGCACATCGAGCAAACTGGGCCTGCAGAGTCAGAAA	872
Db	61		
QY	873	GTCAGGACCTCACTCAGAAGCTGAGGGAGGAGGAAGAAAACTCAAAGCGGTCACCTTAC	932
Db	121	GTACAGGATCTTACTCAGAAGCTGAGGGAAGAGAAGAGCTCAAAGCCATTACTTCC	180
QY	933	AAATCCAAAGGAACCGCCAGAAAGCTGCTCAAGTTAGAAGTGGACTTTCGAACACAAAGGCC	992
Db	181	AAATCCAAAGAGACAGACAGAAATTGCTCAAGTTAGAAGTGGACTTTGAACACAAAGGCT	240
QY	993	TCGAGGTTTTCCAGGAGCACGAAGAGATGAACGCCAAAATTGGCGAATCAAGAATCTCAC	1052
Db	241	TCGAGGTTTTCTCAAGAGCATGAAGAGATGAACGCTAAACTGGCTAATCAAGAGTCTCAC	300
QY	1053	AACCGGCAACTTCGACTCAAACCTGGTTGGCTTATCGCAAAGGATTGAGGAGTGGGAAGAG	1112
Db	301	AATAGGCAACTTAGACTCAAGCTGGTTGGCTTAACCCAAAGAACTCGAGGAGCTAGAAGAG	360
QY	1113	ACCAATAAAAGCCTTCAGAAAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCC	1172
Db	361	ACCAACAAAAATCTGCAGAAAGGCAGAGGAAGAACTTCAAGAAATTAAGAGATAAAAATTGCC	420
QY	1173	AAAGGGGAATGTGGAAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTCGCAAGCGCGTG	1232
Db	421	AAAGGAAATGTGGAAACTCTAGCCTCATGGCAGAAAGTGGAAAAATCTTCGAAAAGCGTGTG	480
QY	1233	CTTGAGATGGAGGGCAAGGATGAAGAGATCAGCAAGACCCGAGGCCCCAGTCCGGGAGCTG	1292
Db	481	CTTGAAATGGAAGGTAAAGATGAGGAGATCTATAAACTGTAATCCCAGTGTAGGGAATTG	540
QY	1293	AAGAAGAAAGCTCCAAGAGGAAGAAACACCACAGCAAGGAACCTTAGACTAGAAGTGGAGAAG	1352
Db	541	AGGAAGAAAGCTGCAAGAGGGAAGAAACACCATAGTAAGGAGCTCAGACTTGAAGTTGAGAAG	600
QY	1353	CTGCAGAAGAGGATGTCTGAGCTGGAGAAGCTTGGAGGAAGCGTTCAGCCGAGTAAGTCG	1412
Db	601	CTACAGAAGAGAAATGTCTGAACCTAGAGAAATTTGAAGAAGCATTTAGCAAGAGTAAATCT	660
QY	1413	GAATGCACCCAGCTCCATCTGAACCTGGAGBAGGAGAAGAACCTTAACCAAGACCTGCTG	1472
Db	661	GAGTGCACCCAGCTACATTTTAAATCTGGAGAAAAGAAAAGAACTTAACCAAGACCTGCTA	720
QY	1473	AACGAGCTGGAGGTGGTCAAGAGTTCGAGTTAAAAGAACTCGAATGCTCCGAGAGTAGACTG	1532
Db	721	AATGAATTGGAGGTGGTCAAGAGTTCGAGTTAAAAGAAATTTGAATGTTCTCGAAAAGTAGATTG	780
QY	1533	GAGAAAGCCGAGTTAAGCCTCAAAGATGACCTTCAAAGCTGAAAGTCTTTCACCTGTGATG	1592
Db	781	GAAAAGGCTGAATTAAGCCTAAAAGATGATCTTACCAAGTTGAAGTCATTTACCGTGATG	840
QY	1593	CTGGTGGATGAGAGGAAAAAATATGATGGAGAAAAATAAGCAAGAAGAGAGCAAGTGGAT	1652
Db	841	CTGGTTGATGAAGGAAAAAATATGATGGAAAAAATAAAACAAGAAGAGAGAAAAAGTGGAT	900
QY	1653	GGGTTGAATAAAAAACTTTTAAGGTGGAGCAGGGAAAAAGTCATGGATGTGACGGAAGAAAGCTA	1712
Db	901	GGACTCAATAAAAAATTTTAAGGTGGAACAAAGGAAAAAGTTATGGATGTAACTGAAAAAACTA	960
QY	1713	ATCGAGGAAGCAAGAAGCTTTTAAAACTCAAATCTGAATGGAGGAAAAAGGAGTACAGT	1772
Db	961	ATTGAAGAAAGTAAGAAACTTTTAAAACTTAAAAATCTGAAATCGAGGAAAAAAGTATACAA	1020
QY	1773	CTGACAAAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGAAGAAAGGTCCTGT	1832
Db	1021	TTGACAAGAGAAAAAGATGAGTTGATAGGCAAAATTGAAAAGTGAAGAAGAAAAATCTCTCT	1080
QY	1833	GAACTGAGCTGCAGTGTAGACTTACTAAAGAAGCGGCTTTGATGGCATAGAGGAGGTAGAA	1892
Db	1081	GAATTAAGCTGCAGTGTTCAGCTTACTTAAAGAAGAGAGACTTTGATGGTATAGAGGAAGTGGAA	1140
QY	1893	AGGAAATAAAACCGAGGTAGGTCGTGCAAGGGGTCTGAGTTACCTGCCCGGAAGACAAAT	1952

Db	1141	AGAGAAATAACAAGAGGAAGGTACGAAAAGGGTCTGAGCTCACCTGCCCGGAAGATAAT	1200
Qy	1953	AAGATCAGAGAACTAACGCTTGAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAG	2012
Db	1201	AAGATTAAAGGAACTAACACTTGAAATTGAGAGACTGAAGAAACGTCTCCAACAATTGGAA	1260
Qy	2013	GTGGTCGAGGGGACTTGATGAAGACCGAGGACGAATATGACAGTTGGAGCGAAGTTC	2072
Db	1261	GTGGTCGAAGGGGATTTGATGAAGACAGAAGATGATATGATCAGCTGGAAACAGAAATTT	1320
Qy	2073	AGAACCGAGCAGGATRAAGCAAACTTCCTCTCCAGCAGCTCGAGGAAATCAACACACAA	2132
Db	1321	AGAACTGAGCAGGATAAGGCTAACTTCCTCTCTCAACAACATAGAGGAGATCAAGCACCAA	1380
Qy	2133	ATGCCAAAGCACAAAGCCATAGAGAAAGGGGAGGCCGTGAGCCAGGAAGCCGAACCTGCGA	2192
Db	1381	ATTGCCAAGAAATAAGCAATAGAGAAAGGTGAGGTTGTGAGCCAGGAAGCTGAACCTGAGA	1440
Qy	2193	CACAGGTTTCGGCTCGAGGAGGCTAAAGTCGTGATTTTACAGGCCGAGGTGACGGCTCTC	2252
Db	1441	CACAGATTCGGTTGGAAGAAGCTAAAGTCGAGACTTAAAGCCGGAGTACAAGCTCTT	1500
Qy	2253	AAGGAGAAGATCCACGAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGAC	2312
Db	1501	AAAGAGAAGATTTCACGAATTAATGAACAAAGAAGATCAGCTTCTCAGCTCCAGGTAGAT	1560
Qy	2313	TATTCCGTCCTTCAGCAAAAGATTTATGGAAGAAGAACTAAAGAACAAGAACATGGGGAGG	2372
Db	1561	TATTCTGTACTTCAACAAAAGATTTATGGAAGAAGAAATAAAGAACAACATGGGGCAG	1620
Qy	2373	GAGTCTCTCAATCTGACCAAGGAGCTAGAGCTTTTCCAAGCGCTACAGCCGAGCTCTCAGG	2432
Db	1621	GAGGTTCTCAATCTGACCAAAAGATTGAGGCTTTTCCAAGCGCTACAGCAGAGCTCTTAGG	1680
Qy	2433	CCGAGTGGGAACGGCCGAAGGATGGTGGACGTGCCCTGTGGCTCCACTGGGGTGACAGCC	2492
Db	1681	CCCAGTGTGAATGGAAGAAAGAAATGGTGGATGT-TCTGTGACGTCAACTGGAGTCCAACT	1739
Qy	2493	GAGCGGTGTGCGGGATGCTGCGGAGGAGGAGACCCCGCTGTGTTCTATTCGCAAAATCC	2552
Db	1740	GATGCAGTCAGCGGTGAAGCAGCAGAGGAAGAAACGCCAGCTGATTTCTACGGAATCC	1799
Qy	2553	TTCAGGAGGAAATCACATCATGAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCCATG	2612
Db	1800	TTCAGGAAGAAATCATATTATGAGTAATCTTCGGCAGGTGGGATTGAAGAAACCCGTG	1859
Qy	2613	GAACGGTCCTCGGTCTCGACAGGTTATCCCCCAGCAGCGCAATGAGCTCACCATGAGGAAG	2672
Db	1860	GAAGATCTTCTGTCTTAGACAGGTATCTCCAGCAGCAAAATGAGCTCATTATGAGAAAG	1919
Qy	2673	TCTTGGATTCTTGGATGAGAAAAGAGAAAAACGGTCTTCCACTCCGCAGGAGAAAGGG	2732
Db	1920	TCTTGGATTCCATGGATGAGAAAGAGGGAACCGGCCCTCCATCACTCAGGAGAAAGGG	1979
Qy	2733	CCAGGCCAAACCCAGGTGCAGGGCACCCCGGGAGCTGGTCTTAGCACCAAGCAGGGC	2792
Db	1980	CCCGAACAAATTCAGTCCAGGGCACCCAGGAGGTAGTCCCTTTCACCAAGCAGGGC	2039
Qy	2793	CAGCCCTACATCCGTGTGACACACAGATCATGAGAAACAGCACTGCCACCCTGGAGATC	2852
Db	2040	CAGCCCTGCATATTCGAGTGACACACAGACCCAGAGAACAGCACTGCCGACTTTGGAGATA	2099
Qy	2853	ACAAGCCCCACATCTGAAGAGTTTCTCTAGTACCACCGTCAATTCCTACCTTAGGCAAC	2912
Db	2100	ACAAGCCCGACATCTGAAGAAATTTTCTTAGTACCAGTGTCAATTCCTACCTTAGGGAAT	2159
Qy	2913	CAGAAACCAAGATAACATTATTCATCACCCTACCCCAATGTCATGTGCAAAAGCCCAAGT	2972
Db	2160	CAGAAACCAAGATAACCATTTATTCATCACCCTACCCCAACGGTATGCCTCAAAACCAAAAGT	2219
Qy	2973	GCAGATCCTACTCTCGGCCCCAGAACGAGCCCATGTCCCTGTACCGATTACTACTATTTC	3032

Db 2220 GGAGATACTACTCTTGGCCAGAACGAGCCATGTCCCCAGTCACAATTACTACATTTTCC 2279

QY 3033 AGAGAGAAAGAGCCCGGAAGGTGGAAGGAGCGCCTTTGCGGACAGGCCTGCATCCCCCATC 3092

Db 2280 AGAGAGAAGACTCCAGAAAGTGGAAGAGGCGCATTTGCAGACAGGCCACATCCCTATT 2339

QY 3093 CAAATCATGACGGTGTCAACATCTGCAGCTCCCACTGAAATCGTGTCTCTCTGAAATCT 3152

Db 2340 CAGATAATGACGGTGTCTACATCAGCAGCACCCAGCTGAGATTGCAGTTTCTCCGGAATCC 2399

QY 3153 CAGGAAGTGCCTATGGGAAGGACTATCTCAAAGTCACCCCGGAAAAACAACACTGTTC 3212

Db 2400 CAGGAAATGCCCATGGGACGGACAATCCTCAAAGTCACCCCGAAGAACAGACTGTTC 2459

QY 3213 GCCCCCGTGGGAAGTACAACTCCAATGCTAATATCATCACCGAAGACATAAAATT 3272

Db 2460 ACTCCAGTACGGAATACAACTCCAATGCCAATATATATAACACAGAGGACATAAAATT 2519

QY 3273 CACATTCACTGGGTCTCAGTTTAAGCGATCTCCTGGGCTGCCGCTGAAGCGGTGAGC 3332

Db 2520 CACATTCACTTAGGGTCTCAGTTTAAACGGTCCCTGGGACTTCAGGTGAAGGAGTCAGT 2579

QY 3333 CCAGTTATCACCGTCCGGCCTGTCAACGTGACAGCGGAGAGGAGTTTCTACAGGCACA 3392

Db 2580 CCAGTTATTACTGTCCGACCACTAAGCTGACAGCCGAAAGGAGGTTTCCACCGCACT 2639

QY 3393 GTCCTTCGCTCTCCAGGAACCACTCTCTTCAAGACCCCGTGTAGCAAGTGACACGAGC 3452

Db 2640 GTCCTTCGCTCTCCAGGAATCACCTCTCCTCACGGCTGGTGCAAGCAAAAGTGACGAGC 2699

QY 3453 ACTATAACTATAACCCCGGTCAACAACGTCAACACAGGAGAACCCCAATCAGTGTGAG 3512

Db 2700 ACTATCACCATTAACCCGGTCAACAACGTCACTGTGAGGAAACCCAGTCAGTGTGAG 2759

QY 3513 CAAGATGGGTCACTCAGCGGCCTACCCGCCATTCCTATGTCAAAGGTATGAAA 3572

Db 2760 CAAGACGGGTCACTCCAGCGGCCTACACCCGCCATTCCTATGTCAAAGGTATGAAA 2819

QY 3573 GCTGGAAAGCCAGTAGTGGCAGCCTCAGGAGCAGGAAATCTGACCAAATTCAGCCTCGA 3632

Db 2820 GCAGGAAAGCCAGTAGTGGCAGCCTCAGGAGCAGGAAATCTGACCAAATTCAGCCTCGA 2879

QY 3633 GCTGAGACTCAGTCTATGAAAATAGAGCTGAAGAAATCTGCAGCCAGCAGCTGCCTCT 3692

Db 2880 GCTGAGACTCAGTCTATGAAAATAGAGCTGAAGAAATCTGCAGCCAGCAGCAGCCTCT 2939

QY 3693 CTTGAGGGGGAGGGCTGAGGGCAGTGGCTAAGGGGTATGTTGTAAGGATGCTACTG 3752

Db 2940 CTCGAGGGGGAGGGCTGAGGGCAGTGGCTAAGGGGTATGTTGTGCAGATGCTACTG 2999

QY 3753 CTGCAGTGGAAACAAACCTTCTCTGTGCCAACCTTCTCTTGTA-CTACTAATTTAAGT 3811

Db 3000 CTGCCGTGAAAGTGAACCTTCTATCTGTTTGCCAGTCTTTACATGTACTAATTTAAGT 3059

QY 3812 TTTAAATATCTTGTATATAAAATAACCAATTTAATAGCCATGACCCCTCCCATTTTGT 3871

Db 3060 TTTAAATATTTGTATTATAAAATAACCAACTAATAACCAT--TTGTCTTTCCCATTTGT 3117

QY 3872 GCATCTGTTTCAATGAGGGGAATAGAAATTAATPAGCAGAAATTTCTGTTGCTGAATGTT 3931

Db 3118 GCATTTGTTTGTATGCTGGGGAACA----AAATTAGCAAAACTATTGCTGTGCTGCTAGA 3173

QY 3932 CTGTTGAAGATGTTGCTCCAGTTTCACTTCTAGCATGTGGCCCAATTCAGGTAG 3991

Db 3174 AGCCAGGGCGTGTGTTCTAGTTCAGTTTTCAGTTTCTAGCAAGTGGACCCCATCAATAGACC 3233

QY 3992 CTCACGAGTTGTGAAGCCCTCAATATCGTCAACCGGAGAGATTGTGAGGACACATTACATA 4051

Db 3234 CATCTGAGC-----CTGTTTCCTCATCAGTTAGATGTGGGACTCAAT--CACA 3280

QY 4052 TGCTCCCAAAGGCTGGCTCCCAATTTTCTAATTGTAAGCCCAACTTTAATAGACTCAGTT 4111

Db 3281 CGCTCTTCAAGTCCGGCTCCCATATTTCCTAATGCAAGCCAAATTTAATGTACCTTGT 3340

QY 4112 CTGTGAT-TTTTTTTTCCAAAAAAAATAATTTTGAATAGGACAGAGTTTAACAGTTGT 4170

Db 3341 CCACAATAATTTTATTAAAAAAATCCTATTACAAAAAATAGACATACITTTAACTATTGT 3400

QY 4171 CATTTTGCACATCAAGCCCATGAGTTTGATATATGCGTTATAAGAAA-AGAATACTTTCA 4229

Db 3401 CATTGCTCTTTTACATCATGAATTTGCTTTATGTGCTGGAAAAACATCACAATAGCTA 3460

QY 4230 GAGCTATCACAGGGTCTCTAAACTTTTGGAAAAACAAAGCCCTAATATGACCTCAGGA 4289

Db 3461 TCACAGGGCCTGGACCTCTAAAAATTTTGCAAAAACAAAGGTTCTAAGATGATTTTCAGGA 3520

QY 4290 AACAAATTTGAACATGAATAAAAAATGGAATGAAATGAACTGTG 4327

Db 3521 AATAATGTGAACATGTAATAAAAAATGGAATGAAATATG 3558

RESULT 7

AAH14449

ID AAH14449 standard; cDNA; 2594 BP.

XX

AC AAH14449;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human cDNA sequence SEQ ID NO:11923.

XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX

OS Homo sapiens.

XX

PN EP1074617-A2.

XX

PD 07-FEB-2001.

XX

PF 28-JUL-2000; 2000EP-00116126.

XX

PR 29-JUL-1999; 99JP-00248036.

PR 27-AUG-1999; 99JP-00300253.

PR 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

XX

PA (HELI-) HELIX RES INST.

XX

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX

DR WPI; 2001-318749/34.

XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

PT

XX

PS Claim 8; SEQ ID NO 11923; 2537pp + Sequence Listing; English.

XX

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 2594 BP; 869 A; 555 C; 580 G; 590 T; 0 U; 0 Other;

Query Match 39.6%; Score 1729.4; DB 4; Length 2594;
Best Local Similarity 80.8%; Pred. No. 0;
Matches 2110; Conservative Q; Mismatches 481; Indels 22; Gaps 7;

QY	1719	GAAAGCAAGAAGCTTTTAAAACTCAAATCTGAAATGGAGGAAAAGGACTACAGTCTGACA	1778
Db	1	GAAAGTAAGAAACCTTTTAAAACTAAAATCTGAAATGGAGGAAAAGTATACAACTTGACA	60
QY	1779	AAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGAAGAAAGTCCCTGTGAACTG	1838
Db	61	AGAGAAAGAGATGAGTTGATAGGCCAAATTGAAAAGTGAAGAAAGAAAATCCTCTGAATTA	120
QY	1839	AGCTGCAGTGTAGACTTACTAAAGACGGCTTGATGCGCATAGAGGAGGTAGAAAGGAA	1898
Db	121	AGCTGCAGTGTGACTTACTAAAGACAGACTTGATGGTATAGAGGAAGTGGAAAGAGAA	180
QY	1899	ATAAACCGAGGTAGTCTGTCGAAGGGGTCTGAGTTCACCTGCCCCGGAAGACAAATAAGATC	1958
Db	181	ATAACAAGAGGAAGGTCACGAAAAGGGTCTGAGCTCACCTGCCCGGAAGATAATAAGATT	240
QY	1959	AGAGAACTAACGCTTGAAATCGAGAGACTGAAGAAACGGCTCCAGCAAGTTGGAGTGGTG	2018
Db	241	AAGGAACTAACACTTGAAATTGAGAGACTGAAGAAACGCTCCCAACAATTGGAAGTGGTC	300
QY	2019	GAGGGGACTTGATGAAGACCGGAGGACGAATATGACCAGTTGGAGCAGAAGTTCAGAAC	2078
Db	301	GAAGGGGATTTGATGAAGACAGAAAGATGAGTATGATCAGCTGGAACAGAAAATTAGAACT	360
QY	2079	GAGCAGGATAAGGCAAACTTCCTCTCCAGCAGCTCGAGGAAATCAAAACACCAAAATGGCC	2138
Db	361	GAGCAGGATAAGGCTTAACCTCTCTCAACAACTAGAGGAGATCAAGCACCAAAATGCC	420
QY	2139	AAGCAAAAGCCATAGAGAAAGGGGAGGCCGTGAGCCAGGAAGCCGAACTGCGACACAGG	2198
Db	421	AAGAATAAGCAATAGAGAAGGTGAGTTGTAGCCAGGAAGCTGAACTGAGACACAGA	480
QY	2199	TTTCGGCTGGAGGAGCTAAAAGTCGTGATTTACAGGCCGAGGTGCAAGGCTCTCAAGGAG	2258
Db	481	TTTCGGTTGGAAGAGCTAAAAGTCGAGACTTAAAAGCCGAAGTACAAGCTCTTAAAGAG	540
QY	2259	AAGATCCACGAGCTGATGAACAAGGAAGACCAAGCTGTCTCAGCTCCAAGTCGACTATT	2318
Db	541	AAGATTCACGAATTAAATGAACAAGAAGATCAGCTTCTCAGCTCCAGGTAGATTATTCT	600
QY	2319	GTCCCTTCAGCAAAAGATTTATGGAAGAAAGAACTAAGAAACAAGAACATGGGGAGGGAGT	2378
Db	601	GTACTTCAACAAGAATTTATGGAAGAAGAAAAAAGAACAACAAACATGGGGCAGGAGGTT	660
QY	2379	CTCAATCTGACCAAGGAGCTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGT	2438
Db	661	CTCAATCTGACCAAGAGTTGGAGCTTTCCAAGCGCTACAGCAGAGCTCTTAGGCCAGT	720
QY	2439	GGGAACGGCCGAAGGATGGTGGACGTGCCCTGTGGCTCCACTGGGGTGCAGACCCGAGCG	2498
Db	721	GTGAATGGAAGAAGATGGTGATGTTCTGTGACGTCAACTGGAGTCCAAACTGATGCA	780
QY	2499	GTGTGCGGGGATGTGCGGAGGAGGAGACCCCGGCTGTGTTCAATTCGCAAAATCCTTCCAG	2558
Db	781	GTACGGGTGAAGCAGAGAGGAAAGAACGCCAGCTGTATTATACGGAAAATCCTTCCAG	840
QY	2559	GAGGAAAATCACATCATGAGTAATCTTCGACAGGTAGGCCCTGAAGAAAACCCATGGAA	2618
Db	841	GAAGAAAATCATATTATGAGTAATCTTCGGCAGGTGGGATTGAAGAAAACCCGTGGAA	900

QY	2619	TCCTCGGTCTTCGACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTGG	2678
Db	901	TCCTTCTGTTCTAGACAGGTATCCTCCAGCAGCAAAATGAGCTCATTATGAGAAAAGTCTTGG	960
QY	2679	ATTCTCTGGATGAGAAAAAGAGAAAACGGTCTTCCACTCCGAGGAGAAAGGGCCAGG	2738
Db	961	ATTCCATGGATGAGAAAAGAGGAAAACGGCCCCCTCCATCACTCAGGAGAAAGGGCCCCGA	1020
QY	2739	CCAAACCCAGGTGCAGGGCACCCCGGGGAGCTGGTCTTAGCACCAAGCAGGGCCAGCCC	2798
Db	1021	ACAAATTCAGTCCAGGGCACCCAGGAGAGGTAGTCTTTACCAAAAGCAGGGCCAGCCC	1080
QY	2799	CTACACATCCGTGTGACACCCAGATCATGAGAACAGCACTGCCACCTGGAGATCACAA	2858
Db	1081	CTGCATATTTCGAGTGACACCCAGACCCAGAACAGCACTGCGACTTTGGAGATAACAAGC	1140
QY	2859	CCACATCTGAAGAGTTTTTCTCTAGTACCACCGTCACTTCTACCTTAGGCAACAGAAA	2918
Db	1141	CCGACATCTGAAGAAATTTTTTCTTAGTACCACCTGTCACTTCTACCTTAGGGAATCAGAAA	1200
QY	2919	CCAGAATAACCATTTATTCATCACCCAATGTCTGTGCAAAAAGCCCCAAAAGTGCAGAT	2978
Db	1201	CCAGAATAAACCATTTATTCATCACCAAAACGTTATGCTCAAAAAACAAAAAAGTGGAGAT	1260
QY	2979	CTACTCTCGGCCAGAACGAGCCATGTCCCCTGTCAAGTATCTACTACTATTTCCAGAGAG	3038
Db	1261	ACTACTCTTGGCCAGAACGAGCCATGTCCCCAGTCAAAATTAATACTATTTTCCAGAGAG	1320
QY	3039	AAGAGCCCCGAAGGTGGAAGGAGCGCCTTTGCCGACAGGCCCTGCATCCCCCATCCAAATC	3098
Db	1321	AAGACTCCAGAAAGTGAAGAGGGCGCATTTGCAGACAGGCCACATCCCCCTATTTCAGATA	1380
QY	3099	ATGACGGTGTCAACATCTGCAGTCCCACTGAAATCGCTGTCTCTCTGAAATCTCAGGAA	3158
Db	1381	ATGACGGTGTCTACATCAGCAGCACCACTGAGATTGCAAGTTCTCCCGAATCCAGGAA	1440
QY	3159	GTGCTATGGGAAGGACTATCCTCAAAGTCACCCCGGAAAAACAACTGTTCCAGCCCCC	3218
Db	1441	ATGCCCATGGGACGACAATCCTCAAAGTCACCCCGAGAAAACAGACTGTTCCCACTCCA	1500
QY	3219	GTGCGGAAGTACAACTCCAATGCTAAATATCATCACACGGAAGACAATAAAATTCACATT	3278
Db	1501	GTACGGAAATACAACTCCAATGCGCAATATATAACACACAGAGGACAATAAAATTCACATT	1560
QY	3279	CACCTGGGTTCTCAGTTTAAAGCATCTCTCTGGGCTGCCGTGAAGCGGTGAGCCAGTT	3338
Db	1561	CACCTAGGGTCTCAGTTTAAACGGTCCCTGGGACTTCAGGTGAAGGAGTCAGTCCAGTT	1620
QY	3339	ATCACCGTCCGGCCTGTCAACGTGCACAGCGGAGAGGAGGTTTCTACAGGCACAGTCTT	3398
Db	1621	ATTACTGTCCGACCAGTAAACGTCACAGCCCGAAAAAGGAGGTTTCACCCGGCACTGTCTT	1680
QY	3399	CGCTCTCCAGGAACCACTCTCTTCAAGACCCCGTGTAGCAAAAGTGACCACTATA	3458
Db	1681	CGCTCTCCAGGAATCACCTCTCTCACGGCCTGGTGAAGCAAAAGTGACGAGCACTATC	1740
QY	3459	ACTATAACCCCGGTCAAAACGTCAATCCACACGAGGAACCAATCAGTGTGAGGACAAGAT	3518
Db	1741	ACCATAAACCCGTCAAAACGTCAATCTGCTCGAGGAACCCAGTCAGTGTGAGGACAAGAC	1800
QY	3519	GGGTCACTCAGCGGCTTACCCCAACCCGCAATTCCTATGTCAAAAGGTATGAAAAGCTGGA	3578
Db	1801	GGGTCACTCCAGCGGCTTACCCCAACCCGCAATTCCTATGTCAAAAGGTATGAAAAGCAGGA	1860
QY	3579	AAGCCAGTAGTGGAGCCTCAGGAGCAGGAAATCTGACCAAAATTCAGCCCTCGAGCTGAG	3638
Db	1861	AAGCCAGTAGTGGAGCCTCCAGGAGCAGGAAATCTGACCAAAATTCGAGCCTCGAGCTGAG	1920
QY	3639	ACTCAGTCTATGAAAAATAGAGCTGAAGAAAATCTGCAGCCAGCAGCACTGCCTCTCTTGG	3698
Db	1921	ACTCAGTCTATGAAAAATAGAGCTGAAGAAAATCTGCAGCCAGCAGCACTCTCTCTCGGA	1980

QY 3699 GGGGGAAGGGCTGAGGCGAGTGGCTAAGGGGGTATGTTTAAGGATGCTACTGCTGCAG 3758
|||||
Db 1981 GGGGGAAGGGCTGAGGCGAGTGGCTAAGGGGGTATGTTGTCAGATGCTACTGCTGCCG 2040
|||||
QY 3759 TGGAAACAAACCTTCCTCTGTCGCCAACCCCTTTCCTTGTA-CTACTAATTAAAGTTTAAA 3817
|||||
Db 2041 TGAAGTGAACCTTCATCTGTTGTGCCAGTTCCTTACATGTACTAATTTAAGTTTAAA 2100
|||||
QY 3818 TATCTTGTTTATAAAATAACCATTTAATAGCCATGCAACCCCTCCCATTTTGTGCATCT 3877
|||||
Db 2101 TATTGTGTTTATAAAATAACCAACTAATAACCAT--TTGCTTTCCCATTTTGTGCATTT 2158
|||||
QY 3878 GTTTCATGTCAGGGGAATAGAATTAATTAGCAGAAATTCTGTTGCTGAATGTTCTGTTG 3937
|||||
Db 2159 GTTTTGATGCTGGGGAACA---AAATTAGCAAAACTATTGCTTGCTGCTAGAACCCAG 2214
|||||
QY 3938 AAGATGTTGGTCCAGTTCAGTTTACTTCTAGCATGTGGCCCCCATTCAGGTTAGCTCAG 3997
|||||
Db 2215 GCGGTGTTTCTAGTTCAGTTTGTCTTCTAGCAAGTGGACCCATCAATAGACCCATCTG 2274
|||||
QY 3998 AGTTGTGAAGCCCTCAATATCGTCACCGGAGAGATTTGAGGACCACATTACATATGCTCC 4057
|||||
Db 2275 AG-----CCTGTTCTCATCAGTTAGATGTGGGACTCAAT--CACACGCTCT 2321
|||||
QY 4058 CAAAGGCTGGCTCCCAATTTTCTAATTGTAAGCCAACTTTAATAGACTCAGTTCTGTGA 4117
|||||
Db 2322 TCAAGTCCGGCTCCCATATTTCTTAATTGCAAGCCAAATTTAATGTACCTTTGTTCCACAA 2381
|||||
QY 4118 T-TTTTTTTTCCAAAAAATAATTTTGAATAGGACAGAGTTTAAACAGTTTGTCAATTT 4176
|||||
Db 2382 TAATTTTATTAAAAAATCCTATTACAAAATAAGACATACCTTTAACTATTGTCAATTG 2441
|||||
QY 4177 GCACTATCAAGCCATGAGTTTGATATATATGGGTTTAAAGAAA-AGAATACTTTCAGAGCTA 4235
|||||
Db 2442 CCTCTTCACATCATGAATTTGCTTTATGTGCTGGAATAACATCACATAGCTATCACAG 2501
|||||
QY 4236 TCACAGGCTCTTAACTTTTGGAAAAACAAAAGCCCTTAATATGACCTCAGGAAAAAAT 4295
|||||
Db 2502 GGCCTGGACCTCTAAAAATTTGCAAAAAACAAAAGGTTCTAAGATGATTTTCAGGAAATAAT 2561
|||||
QY 4296 TTGAACATGAAATAAATGGAAATGAACCTGTGG 4328
|||||
Db 2562 GTGAACATGTAATAAATGGAAATGAAATATGG 2594
|||||

RESULT 8
ACH03975
ID ACH03975 standard; cDNA; 2402 BP.
XX
AC ACH03975;
XX
DT 26-SEP-2003 (first entry)
XX
DE Human cDNA differentially expressed in lung cancer #180.
XX
KW Gene therapy; emphysema; ss; gene; chronic obstructive pulmonary disease;
KW respiratory disorder; lung cancer; asthma; human.
XX
OS Homo sapiens.
XX
PN US2003065157-A1.
XX
PD 03-APR-2003.
XX
PF 04-APR-2002; 2002US-00116802.
XX
PR 04-APR-2001; 2001US-0281593P.
XX
PA (LASE/) LASEK A W.
XX
PI Lasek AW;
XX
DR WPI; 2003-540803/51.

XX New combination comprising cDNAs that are differentially expressed in
PT respiratory disorders, useful for diagnosing or treating respiratory
PT disorders e.g., lung cancer, chronic obstructive pulmonary disease,
PT emphysema or asthma.
XX
PS Claim 1; Page; 39pp; English.
XX
CC The invention relates to a combination comprising cDNAs or their
CC complements that are differentially expressed in respiratory disorder.
CC The combination is useful for preparing a composition for diagnosing or
CC treating respiratory disorders e.g. lung cancer, chronic obstructive
CC pulmonary disease, emphysema or asthma. The present sequence represents
CC human cDNA differentially expressed during lung cancer
XX
SQ Sequence 2402 BP; 769 A; 539 C; 528 G; 566 T; 0 U; 0 Other;

Query Match 35.0%; Score 1527.6; DB 8; Length 2402;
Best Local Similarity 80.4%; Pred. No. 0;
Matches 1894; Conservative 0; Mismatches 439; Indels 23; Gaps 8;

QY 1991 GAAACGGCTCCAGCAGTTGGAGGTGGTGGAGGGGACTTGATGAAGACCGAGGGAATA 2050
|||||
Db 1 GAAACGCTCTCCAACAAATTGGAAGTGGTCGAAGGGGATTGATGAAGACAGAGATGAGTA 60
|||||
QY 2051 TGACCAGTTGGAGCAGAAAGTTCAGAACCGAGCAGGATAAGGCAAACTTCCTCTCCAGCA 2110
|||||
Db 61 TGATCAGCTGGAAACAGAAATTTAGAACTGAGCAGGATAAGGCTAACTTCCTCTCAACA 120
|||||
QY 2111 GCTCGAGGAAATCAAAACCAATGGCCAAAGCAAAAGCCATAGAGAAAGGGAGCCGT 2170
|||||
Db 121 ACTAGAGGAGATCAAGCACCAAAATTGCCAAGATAAAGCAATAGAGAAGGTGAGTTGT 180
|||||
QY 2171 GAGCCAGGAAGCCGAATCGGACACAGGTTTCGGCTGGAGGAGGCTAAAGTCGTGATTT 2230
|||||
Db 181 GAGCCAGGAAGCTGAATGAGACACAGATTTTCGGTTGGAAGAAGCTAAAGTCGAGACTT 240
|||||
QY 2231 ACAGGCCGAGGTGCAAGCTCTCAAGGAGAAGATCCACGAGCTGATGAACAAGGAAGACCA 2290
|||||
Db 241 AAAGCCGAAGTACAAGCTCTTAAGAGAAGATTCACGAATTAATGAACAAGGAAGATCA 300
|||||
QY 2291 GCTGTCTCAGCTCCAAGTCGACTATTTCGGTCTTCAGCAAAAGATTTATGGAAGAAGAAC 2350
|||||
Db 301 GCTTTCTCAGCTCCAGGTAGATTATTCTGTACTTCAACAAAGATTTATGGAAGAAGAAA 360
|||||
QY 2351 TAAGAACAAAGAACATGAGGAGGAGGTCTCAATCTGACCAAGGAGCTAGAGCTTTCCAA 2410
|||||
Db 361 TAAGAACAAACATGAGGAGGAGGTTCTCAATCTGACCAAGAGTTGGAGCTTTCCAA 420
|||||
QY 2411 GCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAAGGATGTTGACGTGCTGT 2470
|||||
Db 421 GCGCTACAGCAGAGCTCTTAGGCCCAGTGTGATGAAGAAGAAATGGTGGATGTTCTGT 480
|||||
QY 2471 GGCCTCCACTGGGGTGCAGACCCGAGCGGTGTGCGGGGATGCTGCGGAGGAGGACACCC 2530
|||||
Db 481 GAGGTCAACTGGAGTCCAAACTGATGAGTCAGTCAGCGGTGAAGCAGCAGAGGAAGAACGCC 540
|||||
QY 2531 GGCTGTGTTTCAATTCGCAAAATCCTTCCAGGAGGAAATCACATCATGAGTAATCTTCGACA 2590
|||||
Db 541 AGCTGTATTTCATACGGAATCCTTCCAGGAAGAAATCATATTATGAGTAATCTTCGGCA 600
|||||
QY 2591 GGTAGGCTGAAGAAACCCCATGGAACGGTCTCGGTCTCTCGACAGGTATCCCCAGCAGC 2650
|||||
Db 601 GGTGGGATTGAAGAAACCCCGTGGAAAGATCTTCTGTTCTAGACAGGTATCTCCAGCAGC 660
|||||
QY 2651 GAATGAGCTCACCATGAGGAAGTCTTGGATTCTTGGATGAGAAAAGAGAAAACGGTCC 2710
|||||
Db 661 AAATGAGCTCACTATGAGAAAGTCTTGGATTCTTCCATGATGAGAAAGAGGGAAAACGGCCC 720
|||||
QY 2711 TTCCACTCCGAGGAGAAAGGGCCCCAGGCCAAACACAGGGTGCAGGGCACCCCCGGGAGCT 2770
|||||
Db 721 CTCCATCACTCAGGAGAAAGGGCCCCGAAACAAATTCCAGTCCAGGGCACCCAGGAGAGGT 780
|||||

XX New Goodpasture antigen binding protein-interacting 90 and 130 kDa
PT polypeptides, useful for diagnosing and/or treating disorders associated
PT with the GIP90/130 polypeptide, such as autoimmune disorders and/or
PT cancer.
XX
PS Claim 48; Page 73-78; 115pp; English.
XX
CC The present full-length cDNA sequence encodes GIP130b, a novel 130 kDa
CC Goodpasture antigen binding protein (GPBP) interacting protein. The cDNA
CC was isolated from a skeletal muscle cDNA library. The DOC/GIP90/130 mRNA
CC family results from a complex diversification mechanism operating on the
CC expression of the GIP90 gene on chromosome 3 (3q12). GIP90/130
CC polypeptides of the invention interact with GPBP and are capable of
CC aggregation. They can be used to modify GPBP-GIP90/130 interactions, to
CC modify GP90/130 aggregation, and to modulate gene expression. The
CC invention provides GIP90/130 polypeptides, portions of them, antibodies,
CC nucleic acid sequences, expression vectors and host cells, as well as
CC methods for detecting GIP90/130 polypeptides and nucleic acids, and
CC methods for treating an autoimmune disease or cancer by modifying the
CC expression or activity of one or more GIP90/130 polypeptides
XX
SQ Sequence 3415 BP; 1293 A; 691 C; 758 G; 673 T; 0 U; 0 Other;

Query Match 18.1%; Score 790.2; DB 8; Length 3415;
Best Local Similarity 55.2%; Pred. No. 8.9e-175;
Matches 1680; Conservative 0; Mismatches 1328; Indels 33; Gaps 6;

QY 286 CTGTGGAGTTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGGAGTTGC 345
DB 202 CAGAAGACCTCTCAAGAGATGACCTGTTATTTCTCTCTCAGCATTTCTGGAGGGAGAACTGC 261

QY 346 AGGCTCGAGAAGATGTCATCCACATGCTGAGGACAGAGAAAACCAAGCCCGAGGTTCTGG 405
DB 262 AGGCTCGAGATGAGGTTCATAGGCATTTTAAAGGCTGAAAAAATGGACCTGGCTTGTCTGG 321

QY 406 AGGCACACTATGGATCTGCAGAAACCTGAGAAAAGTGTCTGGGTCTCTGCACCGAGATGCCA 465
DB 322 AAGCTCAGTATGGGTTTGTCACTCCAAAAAAGGTTAGAGGCTCTCCAGAGAGATGCTT 381

QY 466 TCCTTGCTCAAGAGAAAGTCCATAGGAGAAGACGTCTATGAGAAAACCTATCTCAGAGCTGG 525
DB 382 TTCAAGCGAAATCTACCCCTTGGCAGGAGGACATCTATGAGAAAACCAATGAATCAGTTGG 441

QY 526 ACAGACTGGAGGAAAAGCAGAAAGGAGACGTACCCGCCGATGCTAGAGCAGCTGCTGCTGG 585
DB 442 ACAAGTTGTGGA AAAACATAAAGAATCTTACAGACGAATCCTGGACAGCTTTTAGTGG 501

QY 586 CTGAGAAGTGTACAGGCGCACCGTGTACGAGCTGGAGAACGAGAAGCACACACTG 645
DB 502 CAGAAAATCCCATAGGCAACCATATTGGAGTTGGAGGAAGAAAGAGAAAACATAAAG 561

QY 646 ACTACATGAACAAGCGCAGCTTCACCAACCTGCTGGAGCAGGAGCGAGAGAGTTGA 705
DB 562 AATACATGGAGAAGAGTGAATTCATATGCCTACTAGAACAGGAATGTGAAGATTA 621

QY 706 AAAAGCTCCTTGAACAAGAAAAAGCTTACCAGCCCGCAAAAGAAAAGGAAAACGCTAAGC 765
DB 622 AGAAGCTAATTGATCAAGAAATCAAGTCTCAGGAGGAGAAGGAGCAAGAAAAGGAGAAAA 681

QY 766 GGCTCAACAAAACCTTCGAGATGAGCTTGTGAAGCTCAAGTCTTCGCCCTCATGTTGGTGG 825
DB 682 GGGTCACCAACCTTGAAGAGGAGCTGACCAAGCTGAAGTCTTTTGTCTTTGATGGTGGTGG 741

QY 826 ACGAGAGGCAGATGCACATCGAGCAACTGGGCCCTGCAGAGTCAGAAAAGTCCAGGACCTCA 885
DB 742 ATGAACAGCAAAGGCTGACGGCACAGCTCACCCCTTCAAAAGACAGAAAATCCAAGAGCTGA 801

QY 886 CTCAGAAGCTGAGGGAGGAGGAAGAAAACCTCAAAGCGGTCACTTACAAATCCAAGGAAG 945
DB 802 CCACAAATGCAAAGGAAACACATACCAAACCTAGCCCTTGTGTAAGCCAGAGTTCAGGAGG 861

QY 946 ACCGCCAGAAGCTGCTCAAGTTAGAAGTGGACTTCGAACACAAAGGCCCTCGAGGTTTCCC 1005

DB 862 AAGAGCAGAAGGCAACAGACTAGAGAAGGAACCTGCAAAACGCAGACCACAAAAGTTTCACC 921
QY 1006 AGGAGCACGAAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCACAAACCGCAACTTC 1065
DB 922 AAGACCAAGACACAATATATGGCGAAGCTCACCAATGAGGACAGTCAAAATCGCCAGCTTC 981

QY 1066 GACTCAAACTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGAAGAGACCAATAAAGCC 1125
DB 982 AACAAAAGCTGGCAGCACTCAGCCGGCAGATTGATGAGTTAGAAGAGACAAACAGGTCTT 1041

QY 1126 TTCAGAAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAATTGCCAAAGGGGAATGTG 1185
DB 1042 TACGAAAAGCAGAAAGAGGAGCTGCAAGATATAAAGAAAATAATCAGTAAGGGAGATATG 1101

QY 1186 GAAACTCCAGTCTATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTGTGATGAGATGGAGG 1245
DB 1102 GAAACGCTGGTATCATGGCTGAAGTGAAGAGCTCAGGAAAACGTGTGCTAGATATGGAAG 1161

QY 1246 GCAAGGATGAAGAGATCACGAAGACCGAGGCCAGTCCAGTCCGGGAGCTGAAGAAGAGCTCC 1305
DB 1162 GAAAAGATGAAGAGCTCATAAAAATGGAGGAGCAGTGCAGAGATCTCAATAAGAGGCTTG 1221

QY 1306 AAGAGGAAGAACACCCACAGCAAGGAACCTTAGACTAGAAAGTGGAGAAGCTGCAGAAAGGA 1365
DB 1222 AAAGGGAGACGTTACAGAGTAAGACTTTTAAACTAGAGGTTGAAAAAAGCTCAGTAAAAAG 1281

QY 1366 TGTCTGAGCTGGAGAAAGCTGGAGGAAGCGTTTCAAGCGGAGTAAGTCGGAATGCACCCAGC 1425
DB 1282 TTATGGCTCTGGAAAAAGTTAGAAAGACGCTTTCAACAAAAGCAAAACAAGATGCTACTCTC 1341

QY 1426 TCCATCTGAACCTGGAGAAAGGAGAAAGAACCTAACCAAGACCTGCTGAACGAGCTGGAGG 1485
DB 1342 TGAATGCAATTTAGAAAAAGAAAGGATGACCACAAAAGCAGTTGTCTCAAGAACTGGAGA 1401

QY 1486 TGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAAGGCCGAGT 1545
DB 1402 GTTTAAAAGTAAGGATCAAGAGCTAGAAAGCCATTGAAAGTCGGCTAGAAAAGACAGAAAT 1461

QY 1546 TAAGCCTCAAAGATGACCTTACAAAGCTGAAGTCTTCACTGTGATGCTGTTGGATGAGA 1605
DB 1462 TCACCTTAAAGAGGATTTAACTAAACTGAAAACATTAACCTGTGATGTTTGTAGATGAAC 1521

QY 1606 GGAATAATATGATGGAGAAAATAAAGCAAGAGAGAGGAAAAGTGGATGGTTGAATAAAA 1665
DB 1522 GGAATAACATGATGAAAAAATTAAAGAAAACCTGAAGATAAATTACAAGCTGCTTCTCTC 1581

QY 1666 ACTTTAAGGTGGAGAGGGAAAAAGTCAATGGATGTGACGGAAAAAGCTAATCGAGGAAAGCA 1725
DB 1582 AGCTTCAAGTGGAGCAAAAATAAAGTAACAACAGTTACTGAGAAGTTAATTTGAGGAAACTA 1641

QY 1726 AGAAGCTTTTAAACTCAAAATCTGAAAATGGAGGAAAAAGGAGTACAGTCTGACAAAAGGAGA 1785
DB 1642 AAAGGGCGCTCAAGTCCAAAACCGATGTAGAAGAAAAAGATGTACAGCGTAACCAAGGAGA 1701

QY 1786 GGGATGAGCTGATGGTTAAACTGAGGAGCGGAAGAAAGAGTCCCTGTGAACCTGAGCTGCA 1845
DB 1702 GAGATGATTTAAAAAACAAATTGAAAGCGGGAAGAGAAAGGAAATGATCTCTCTGTCAA 1761

QY 1846 GTGTAGACTTACTAAAGAACGGCTTGTATGGCATAGAGGAGGTAGAAAAGGGAATATAACC 1905
DB 1762 GAGTTAATATGTTGAAAAAATAGGCTTCAATCATTTGGAAGCAATTGAGAAAAGATTTCTCTAA 1821

QY 1906 GAGGTAGGTCGTGCAAGGGGTCTGAGTTCCAC-----TGCCCCGGAAGACAATA 1953
DB 1822 AAAACAAATTTAAATCAAGACTCTGGGAAATCCCAACACAGCATTTACACCAAGAAAAACAATA 1881

QY 1954 AGATCAGAGAACTAACGCTTGAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGG 2013
DB 1882 AGATTAAGGAGCTCTCTCAAGAAGTGGAAAGACTGAAACTGAAGCTAAAGGCACATGAAAG 1941

QY 2014 TGGTGGAGGGGACTTGATGAAGACCGGAGGACGCAATATGACCAGTTGGAGCAGAGTTCA 2073

Db 1942 CCATTGAGGATGACCTCATGAAAAACAGAAAGATGAATATGAGACTCTTAGAACGAAGGTATG 2001

QY 2074 GAACCGAGCAGGATAAGGCAAACTTCCTCTCTCCAGCAGCTCGAGGAATCAAAACCCAAA 2133

Db 2002 CTAATGAACGAGACAAAGCTCAATTTTATCTAAAGACTAGAACATGTTAAATGGAAC 2061

QY 2134 TGGCCAAGCACAAAGCCATAGAGAAAAGGGGAGGCCGTGAGCCAGGAAGCCGAACTGCGAC 2193

Db 2062 TTGCTAAGTACAAAGTTAGCAGAAAAGACAGAGACC---AGCCATGAACAAATGGCTTTTCA 2118

QY 2194 ACAGGTTTCGGCTGGAGGAGGCTAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCA 2253

Db 2119 AAAGGCTTCAAGAAGAAGCTAAGTCAGGGCACCTCTCAAGAGAAGTGGATGCATTAA 2178

QY 2254 AGGAGAAGATCCACGAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACT 2313

Db 2179 AAGAGAAAATTCATGAATACATGGCAACTGAAGACCTTAATATGTCACTCCAGGGAGATC 2238

QY 2314 ATTGGTCCTTCAGCAAAAGATTTATGGAAGAAGAACTAAAGAACAAAGAACATGGGAGGG 2373

Db 2239 ACTCAGTCCTGCAAAAAAAACTAAATCAACAAGAAAACAGGAACAGAGATTTAGGAAGAG 2298

QY 2374 AGGTCCTCAATCTGACCAAGGAGCTAGAGCTTTCRAAGCGCTACAGCCGAGCTCTCAGGC 2433

Db 2299 AGATTCAAAACCTCACTAAGGAGTTAGAGAGGTACCGGCATTTTCAGTAAAGCCTCAGGC 2358

QY 2434 CGAGTGGGAACGGCCGAAGGATGGTGGACGTGCCTGTGGCCTCCACTGGGGTGCAGACCG 2493

Db 2359 CTAGTCTCAATGGAAGAAGAATTTCCGATCCTCAAGTATTTTCTAAAGAAGTTTCAGACAG 2418

QY 2494 AGGCGTGTGCGGGGATGCTGCGGAGGAGGAGACCCCGGCTGTGTTCAATTCGCAATCCT 2553

Db 2419 AAGCAGTAGACAAATGAACCACTGATTAACAAGAGCCTCATTCCTCTGGAACGTGCACTCA 2478

QY 2554 TCCAGGAGGAAAATCACATCATGAGTAATCTTCGACAGGTAGGCCCTGAAGAAACCCATGG 2613

Db 2479 TCAATGTCAGTTATATGAGGAGAGTGAGAATCAAGACGAGGACCCCTAATGAT-----G 2532

QY 2614 AACGFTCCTCGTCTCGACAGGTTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGT 2673

Db 2533 AGGGATCTGTGCTGTCTTCAAATGCAGCCAGTCTACTCCATGTCTCTGTTAACAGAAAGC 2592

QY 2674 CTTGGATTCCTTGGATGAGAAAAAGAGAAAACGGTCTCTCCACTCCGACGAGAAAGGC 2733

Db 2593 TATGGATTCCTGGATGAAATCCAAGGAGGCCATCTTCAGAATGGAATAATGCAA--A 2649

QY 2734 CCAGGCCAAACCAAGGTGCAGGGCACCCCGGGAGCTGGTCTTAGCACCAAAAGCAGGCCC 2793

Db 2650 CTAACCCCAATGCCAACTTTGTGCAACCTGGAGATCTAGTCTCTAAGCCACACACCTGGGC 2709

QY 2794 AGCCCTACACATCCGTGTGACACCAAGATCATGAGAACAGCACTGCCACCCCTGGAGATCA 2853

Db 2710 AGCCACTTCATATAAAGGTTACTCCAGACCATGTACAAAACACAGCCACTCTTGAATCA 2769

QY 2854 CAAGCCCCACATCTGAAG-----AGTTTTCCTAGTACCACCGTCATTCCTACCTTAG 2907

Db 2770 CAAGTCCAACCCACAGAGAGTCTCTCACTCTTACACGAGTACTGCAGTGATACCGAACTGTG 2829

QY 2908 GCAACCCAGAAAACCAAGAATAACCAATTATTCATCACCCAATGTCAITFCGCAAAAGGCCA 2967

Db 2830 GCACGCCAAAGCAAAGGATAAACCATCTCTCCAAAACGGCTCCATAACACCAAGTAAGTCCA 2889

QY 2968 AAA---GTGCAGATCCTACTCTCGGCCAGAACGAGCCATGTCCCTGTTCAGGTTACTA 3024

Db 2890 AAACCTCTACCGAAGACCTCATGAATTTAGAACCAAGGCATGTCCCCCAATTACCATGGCAA 2949

QY 3025 CTATTTCCAGAGAGAGAGCCCGGAAGTGGGAAGGAGCGCCTTTGCCGACAGGCGCTGCAT 3084

Db 2950 CCTTTGCCAGAGACACAGACCCACAGAGTCTTGTGGTTCTCTAACTCCAGAAAAGGACAATGT 3009

QY 3085 CCCCCATCCAAATCATGACGGTGTCAACATCTGAGTCCCACTGAAATCGTGTCTCTC 3144

Db 3010 CCCCTATTACAGGTTTGGCTGTGACTGGTTCAGCTAGCTCTCCTGACAGGGACGCTCCC 3069

QY 3145 CTGAATCTCAGGAAGTGCCTATGGAAGGACTATCTCTCAAAAGTCACCCCGGAAAAACAAA 3204

Db 3070 CAGAACCAACAGAAATCAGTGCCAAGCATGCGATATTCAGAGTCTCCCCAGACCCGCACT 3129

QY 3205 CTGTTCCAGCCCCCGTGGCGGAAGTACAACTCCAATGCTAATATCATCACCACGGAAGACA 3264

Db 3130 CATCATGGCAGTTTCAGCGTTTCAACACAGCAATAGCTCAAGTGTGATAACTACTGAGGATA 3189

QY 3265 ATAAATTCACATTCACCTGGGTTCTCAGTTTAAAGCGATCT 3305

Db 3190 ATAAATCCACATTCACCTTAGGAAGTCCTTACATGCAAGCT 3230

RESULT 10
ACC83930
ID ACC83930 standard; cDNA; 3416 BP.
XX
AC ACC83930;
DT 22-SEP-2003 (first entry)
XX
GPBP-interacting protein GIP130c coding sequence.
DE
XX
KW Goodpasture antigen binding protein interacting protein; GPBP; GIP130c;
human; transcription factor; autoimmune disease; cancer;
immunosuppressive; cytostatic; chromosome 3q12; gene; ss.
KW
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 12..3410
FT /*tag= a
FT /product= "GIP130c"
XX
PN WO2003048193-A2.
XX
PD 12-JUN-2003.
XX
PF 05-DEC-2002; 2002WO-EP013802.
XX
PR 07-DEC-2001; 2001US-0338287P.
PR 20-MAY-2002; 2002US-0382004P.
XX
PA (SAUS/) SAUS J.
PA (REVE/) REVERT-ROS F.
XX
PI Saus J, Revert-Ros F;
XX
DR WPI; 2003-505281/47.
DR P-PSDB; ABR62253.
XX
PT New Goodpasture antigen binding protein-interacting 90 and 130 kDa
polypeptides, useful for diagnosing and/or treating disorders associated
with the GIP90/130 polypeptide, such as autoimmune disorders and/or
cancer.
XX
PS Claim 48; Page 83-88; 115pp; English.
XX
CC The present full-length cDNA sequence encodes GIP130c, a novel 130 kDa
Goodpasture antigen binding protein (GPBP) interacting protein. The cDNA
was isolated from a skeletal muscle cDNA library. The DOC/GIP90/130 mRNA
family results from a complex diversification mechanism operating on the
expression of the GIP90 gene on chromosome 3 (3q12). GIP90/130
polypeptides of the invention interact with GPBP and are capable of
aggregation. They can be used to modify GPBP-GIP90/130 interactions, to
modify GP90/130 aggregation, and to modulate gene expression. The
invention provides GIP90/130 polypeptides, portions of them, antibodies,
nucleic acid sequences, expression vectors and host cells, as well as
methods for detecting GIP90/130 polypeptides and nucleic acids, and
methods for treating an autoimmune disease or cancer by modifying the
expression or activity of one or more GIP90/130 polypeptides

SQ	Sequence 3416 BP; 1293 A; 692 C; 758 G; 673 T; 0 U; 0 Other;	
	Query Match 18.1%; Score 788.6; DB 8; Length 3416;	
	Best Local Similarity 55.2%; Pred. No. 2.1e-174;	
	Matches 1679; Conservative 0; Mismatches 1329; Indels 33; Gaps 6;	
QY	286 CTGTGGAGTTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGAGTTGC	345
Db		
QY	346 AGGCTCGAGAAAGATGTCATCCACATGCTGAGGACAGAGAAAAACCAAGCCGAGGTTCTGG	405
Db		
QY	406 AGGCACACTATGGATCTGCAGAACCTTGAGAAAAGTGCCTTCGGGTCTGCAACCGAGATGCCA	465
Db		
QY	319 AAGCTCAGTATGGGTTTGTCACCTCCAAAAAGGTGTAGAGGCTCTCCAGAGAGATGCTT	378
QY	466 TCCTTGCTCAAGAGAAGTCCATAGGAGAAGACGCTCTATGAGAAAACCTATCTCAGAGCTGG	525
Db		
QY	526 ACAGACTGGAGGAAAAGCAGAAGGAGACGTACCGCCGCGATGCTAGAGCAGCTGCTGCTGG	585
Db		
QY	586 CTGAGAAAGTGTACAGGCGCACCGTGTACGAGCTGGAGAACGAGAGCAACAGCACACTG	645
Db		
QY	646 ACTACATGAACAAGAGCGGACGACTTCACCAACCTGCTGGAGCAGGAGCGGAGAGGTTGA	705
Db		
QY	706 ANAAGCTCCTTGAACAAGAAAAGCTTACCAAGCCCGCAAAGAAAAGGAAAACGCTAAGC	765
Db		
QY	766 GGCTCAACAAAACCTTCGAGATGAGCTTGTGAAGCTCAAGTCTCAAGTCTCGCCCTCATGTTGGTGG	825
Db		
QY	826 ACGAGAGGCAGATGCACATCGAGCAACTGGGCCCTGCAGAGTCAGAAAAGTCCAGGACCTCA	885
Db		
QY	886 CTCAGAAAGCTGAGGGAGGAGGAAGAAAACCTCAAAGCGGTCACTTACAAATCCCAAGGAAG	945
Db		
QY	946 ACCGCCAGAAGCTGCTCAAGTTAGAAAGTGGACTTCGAAACACAAGGCCTCGAGGTTTCCC	1005
Db		
QY	1006 AGGAGCAGAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCAACACCGGCAACTTC	1065
Db		
QY	1066 GACTCAAACTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCC	1125
Db		
QY	1126 TTCAGAAAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGTG	1185
Db		
QY	1186 GAAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTGTGAGATGGAGG	1245
Db		
QY	1246 GCAAGGATGAAGAGATCAAGAACCGAGGCCAGTGCAGGAGCTGAAGAAAGAGCTCC	1305
Db		

QY	1306 AAGAGGAAGAACACCACAGCAAGGAACCTTAGACTAGAGTGGAGAAAGCTGCAGAAGAGGA	1365
Db		
QY	1366 TGTCTGAGCTGGAGAAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGCACCCAGC	1425
Db		
QY	1425 TCCATCTGAACCTGGAGAGGAGGAAGAACCTTAACCAAGAACCTGCTGAACGAGCTGGAGG	1485
Db		
QY	1486 TGGTCAAGAGTCTGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAAGCCGAGT	1545
Db		
QY	1546 TAAGCCTCAAAGATGACCTTACAAAAGCTGAAAGTCCCTTCACCTGTGATGCTGGTGATGAGA	1605
Db		
QY	1605 GAAAAAATATGATGGAGAAAAATAAAGCAAGAAGAGAGGAAAAAGTGGATGGTTGAATAAAA	1665
Db		
QY	1666 ACTTTAAGTGGAGCGGAAAAAGTCAATGGATGTGACGGAAGAAAGCTAATCGAGGAAAGCA	1725
Db		
QY	1726 AGAAGCTTTTAAAACTCAAATCTGAAAATGGAGGAAAAAGGAGTACAGTCTGACAAAAGGAGA	1785
Db		
QY	1786 GGGATGAGCTGATGGGTAAAACTGAGGAGCGAAGAAAGAAAGGTCTCTGTGAACCTGAGCTGCA	1845
Db		
QY	1845 GTGTAGACTTACTAAAGAACGGCTTGTATGGCATAGAGGAGGTAGAAAGGGAATAAACC	1905
Db		
QY	1906 GAGGTAGTCTGTGCAAGGGGTCTGAGTTCCACC-----TGCCCGGAAGACAATA	1953
Db		
QY	1954 AGATCAGAGAACTAACCGTTGAAATCGAGAGACTGAAGAAAAACGGCTCCAGCAGTTGGAGG	2013
Db		
QY	2014 TGGTGGAGGGGACTTGTATGAAGACCGAGGACGCAATATGACCAGTTGGAGCAGAAAGTTCA	2073
Db		
QY	2074 GAACCGAGCAGGATAAGGCAAACTTCTCTCTCCAGCAGCTCGAGGAAATCAAAACACCAAA	2133
Db		
QY	2134 TGGCCAAAGCACAAAGCCATAGAGAAAAGGGGAGGCCGTGAGCCAGGAAGCCGAACCTGCGAC	2193
Db		
QY	2194 ACAGGTTTCGGCTGGAGGAGGCTAAAAGTCGTGATTTTACAGGCCGAGGTGCAGGCTCTCA	2253
Db		
QY	2254 AGGAGAAGATCCACGAGCTGATGAACAAGGAAGACACAGCTGTCTCAGCTCCAAAGTCGACT	2313
Db		
QY	2314 ATTCGGTCTCTCAGCAAAAGATTTTATGGAAGAAGAAACTAAGAAACAAGACATGGGGAGGG	2373
Db		
QY	2373 ACTCAGTCTGCAAAAAAATACTAAATCAACAAGAAAAACAGGAACAGAGATTTTAGGAAGAG	2395
Db		

PA	(REVE/) REVERT-ROS F.	
XX		
PI	Saus J, Revert-Ros F;	
XX		
DR	WPI; 2003-505281/47.	
DR	P-PSDB; ABR62250.	
XX		
PT	New Goodpasture antigen binding protein-interacting 90 and 130 kDa	
PT	polypeptides, useful for diagnosing and/or treating disorders associated	
PT	with the GIP90/130 polypeptide, such as autoimmune disorders and/or	
PT	cancer.	
XX		
PS	Claim 48; Page 54-59; 115pp; English.	
XX		
CC	The present full-length cDNA sequence encodes GIP90, a novel 90 kDa	
CC	Goodpasture antigen binding protein (GPBP) interacting protein. The cDNA	
CC	was isolated following a yeast two-hybrid screen for GPBP interactive	
CC	proteins and use of a partial clone to screen a human skeletal muscle	
CC	cDNA library. The GIP90 gene was localised to chromosome 3 (3q12).	
CC	Comparison of GIP90 cDNA and genomic sequences revealed the existence of	
CC	an adenine at position 2720 in the cDNA that was not present in GIP90	
CC	genomic DNA. Another cDNA fragment was isolated (see ACC83928) that	
CC	encoded a 130 kDa protein (GIP130a) and resulted from faithful	
CC	transcription and translation of the GIP90 genomic sequence (i.e. no	
CC	A2720). GIP90/130 polypeptides interact with GPBP and are capable of	
CC	aggregation. They can be used to modify GPBP-GIP90/130 interactions, to	
CC	modify GP90/130 aggregation, and to modulate gene expression. The	
CC	invention provides GIP90/130 polypeptides, portions of them, antibodies,	
CC	nucleic acid sequences, expression vectors and host cells, as well as	
CC	methods for detecting GIP90/130 polypeptides and nucleic acids, and	
CC	methods for treating an autoimmune disease or cancer by modifying the	
CC	expression or activity of one or more GIP90/130 polypeptides	
XX		
SQ	Sequence 3998 BP; 1448 A; 833 C; 895 G; 822 T; 0 U; 0 Other;	
	Query Match 17.8%; Score 777.6; DB 8; Length 3998;	
	Best Local Similarity 55.2%; Pred. No. 8.7e-172;	
	Matches 1679; Conservative 0; Mismatches 1329; Indels 34; Gaps 7;	
QY	286 CTGTGGAGTTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGAGTTGC	345
Db	663 CAGAAGACCTCTCAAGAGATGACCTGTTATTTCTCTCAGCATCTCTGGAGGAGAACTGC	722
QY	346 AGGCTCGAAGAAGATGTCTATCCATGTCTGAGGACAGAGAAACCAAGCCCGAGTTCTGG	405
Db	723 AGGCTCGAGATGAGGTCTAGGCATTTTAAAGGCTGAAAAAATGGACCTGGCTTGTCTGG	782
QY	406 AGGCACACTATGGATCTGCAGAAACCTGAGAAAGTGTCTCGGCTCTGCACCGAGATGCCA	465
Db	783 AAGCTCAGTATGGGTTTGTCACTCCAAAAAAGGTGTTAGAGGCTCTCCAGAGAGATGCTT	842
QY	466 TCCTTGCTCAAGAGAAGTCCATAGGAGAGACCTCTATGAGAAACCTATCTCAGAGCTGG	525
Db	843 TTCAAGCGAAATCTACCCCTTGGCAGGAGACATCTATGAGAAACCAATGAATGATGTGG	902
QY	526 ACAGACTGGAGGAAAAAGCAGAGGAGACGTACCGCCGCATGTAGAGCAGCTGCTGCTGG	585
Db	903 ACAAAGTTGTGAAAAACATAAAGATCTTACAGACGAATCCTGGGACAGCTTTTAGTGG	962
QY	586 CTGAGAAAGTGTACAGGGCGCACCGTGTACGAGCTGGAGAACGAGAGACACAAGCACA	645
Db	963 CAGAAAAATCCCGTAGGCAACCATATTTGGAGTTGGAGGAAGAGAAACATATAAG	1022
QY	646 ACTACATGAACAAGAGCGACGACTTCAACCACTCTGGAGCAGAGCGGAGAGGTTGA	705
Db	1023 AATACATGGAGAAGTGTATGAATTCATATGCCTACTAGAACAGGAATGTGAAAGATTA	1082
QY	706 AAAAGCTCCTTGAAACAAGAAAAAGCTTACCAAGCCCGCAAGAAAAACGCTAAGC	765
Db	1083 AGAAGCTAATTGATCAAGAAATCAAGTCTCAGGAGGAGGAAGCAAGAAAGGAGAAA	1142
QY	766 GGCTCAACAAACCTTCGAGATGAGCTGTGAAAGCTCAAGTCTTCCGCTCATGTTGTGG	825

Db 1143 GGGTCAACCACTGAAAGAGGAGCTGACCAAGCTGAAAGTCTTTTGTGCTTTGTGTTGGTGG 1202

QY 826 ACGAGAGGCAGATGCACATCGAGCAACTGGGCCCTGCAGAGTCAGAAAAGTCCAGGACCTCA 885

Db 1203 ATGAACAGCAAGGCTGACGGCACAGCTCACCCCTTCAAGACAGAAAATCCAAGAGCTGA 1262

QY 886 CTCAGAAAGCTGAGGAGGAGGAAGAAAACCTCAAAGCGGTCACTTACAAATCCAAGGAAG 945

Db 1263 CCACAAATGCAAGGAAAACACATACCAAACTAGCCCTTGTGAAGCCAGAGTTCAGGAGG 1322

QY 946 ACCGCCAGAAGCTGCTCAAGTTAGAAAGTGGACTTCGAACACAAAGGCCCTCGAGGTTTCCC 1005

Db 1323 AAGAGCAGAAGGCCAACACAGACTAGAGAAGGAACTGCAAAACGCAGACCACAAAGTTTCACC 1382

QY 1006 AGGAGCACGAAGAGATGAACGCCAAATTTGGCGAATCAAGAAATCTCACAAACCGSAACTTC 1065

Db 1383 AAGACCAAGACACAATTTATGGCGAAGCTCACCAATGAGGACAGTCAAAATCGCCAGCTTC 1442

QY 1066 GACTCAAACTGGTTGCTTATCGCAAAGGATTTGAGGAGCTGGAAGAGACCAATAAAAGCC 1125

Db 1443 AACAAAAGCTGGCAGCACTCAGCCGGCAGATTTGATGAGTTAGAAGAGACAAACAGGTCTT 1502

QY 1126 TTCAGAAAGGCAGAGGAAAGACTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGTG 1185

Db 1503 TACGAAAAGCAGAAGAGAGCTGCAAGATATAAAAGAAAAAATCAGTAAGGGAATAATG 1562

QY 1186 GAAACTCCAGTCTCATGCGGAAGTGGAGAGTCTGCGCAAGCGCGTGTCTTGAGATGGAGG 1245

Db 1563 GAAACGCTGGTATCATGCTGAAGTGGAAGAGCTCAGGAAACGTGTGCTAGATATGGAAG 1622

QY 1246 GCAAGGATGAAGAGATCACGAAGACCGAGGCCAGTCCCGGGAGCTGAAGAAAGCTCC 1305

Db 1623 GGAAGATGAAGAGCTCATAAAATGGAGGAGCAGTCAGAGATCTCAATAAGAGGCTTG 1682

QY 1306 AAGAGGAAGAACACCAACAGCAAGAACTTAGACTAGAAGTGGAGAAGCTGCAGAAGAGGA 1365

Db 1683 AAAGGGAGACGTTACAGAGTAAAGACTTTTAACTAGAGTTGAAAAACTCAGTAAAAGAA 1742

QY 1366 TGTCTGAGCTGGAGAGCTGGAGGAAGCGTTTACCGCGAGTAAAGTCGGAATGCACCCAGC 1425

Db 1743 TTATGGCTCTGGAAGAGTTAGAAGACGCTTTCAACAAAAGCAAAACAAAGATGCTACTCTC 1802

QY 1426 TCCATCTGAACCTGGAGAAGGAGAAGAACCTTAACCAAAAGACCTGCTGAACGAGCTGGAGG 1485

Db 1803 TGAATGCAATTTAGAAAAAGAAAAGGATGACCACAAAAGCAGTTGTCTCAAGAACTGGAGA 1862

QY 1486 TGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAGGCCGAGT 1545

Db 1863 GTTTAAAGATAAGGATCAAGAGCTAGAAGCCATTGAAAAGTCGGCTAGAAAAGACAGAAT 1922

QY 1546 TAAGCCCTCAAAGATACCTTACAAAGCTGAAGTCTTCACTGTGATGCTGGTGGATGAGA 1605

Db 1923 TCACCTTAAAAGAGGATTTAACTAAACTGAAACTGAAACATAAATTAAGTGTGATGTTGTAGATGAAC 1982

QY 1606 GGAATAATATGATGGAGAAAATAAAGCAAGAGAGAGGAAAAGTGGATGGTTGAATAAAA 1665

Db 1983 GGAATAACATGAGTGAAAAAATTAAAGAAAACCTGAAGATAAAATTACAAGCTGCTTCTCTC 2042

QY 1666 ACTTTAAGGTGGAGCAGGAAAAAGTCAATGATGTGACGGAAGAAAGCTAATCGAGGAAAGCA 1725

Db 2043 AGCTTCAAGTGGAGCAAAAATAAAGTAAACACAGTTACTGAGAAGTTAATTGAGGAAAACTA 2102

QY 1726 AGAAGCTTTTAAACTCAAACTCTGAAATGGAGGAAAAAGGAGTACAGTCTGACAAAGGAGA 1785

Db 2103 AAAGGGCGCTCAAGTCCAAAACCGATGTAGAAGAAAAGATGTACAGCGTAAACCAAGGAGA 2162

QY 1786 GGGATGAGCTGATGGGTAAACTGAGGAGCGGAAGAAAGGTCCTGTGAACTGAGCTGCA 1845

Db 2163 GAGATGATTTAAAAAACAAAATTGAAAGCGGAAGAGAGAAAAGGAAATGATCTCCTGTCAA 2222

QY 1846 GTGTAGACTTACTAAAGAACGGCTTGATGGCATAGAGGAGGTAGAAAAGGGAATAAAACC 1905

Db 2223 GAGTTAATATGTTGAAAAATAGGCTTCAATCATTTGGAAGCAATTTGAGAAAGATTTCCCTAA 2282

QY 1906 GAGGTAGGTCGTGCAAGGGGTCTGAGTTTACC-----TGCCCGGAAGACAATA 1953

Db 2283 AAAACAAATTAATCAAGACTCTGGGAAATCCACAACAGCATTTACACCAAGAAAACAATA 2342

QY 1954 AGATCAGAGAACTAAACGCTTGAAATCGAGAGACTGAAGAAAACGGCTCCAGCAGTTGGAGG 2013

Db 2343 AGATTAAGGAGCTCTCTCAAGAAAGTGAAGAGACTGAAACTGAAGCTAAAGGACATGAAAG 2402

QY 2014 TGGTGGAGGGGACTTGTATGAAGACCCGAGGACGAATATGACCAGTTGGAGCAGAAATTCA 2073

Db 2403 CCATTGAGGATGACCTCATGAAAAACAGAAAGATGAATATGAGACTCTAGAACGAAGGTATG 2462

QY 2074 GAAACGAGCAGGATAAGGCAAACTTCTCTCTCCAGCAGCTCGAGGAAATCAAAACACCAAA 2133

Db 2463 CTAATGAACGAGACAAAAGCTCAATTTTATCTAAAGAGCTAGAACATGTTAAAAATGGAAC 2522

QY 2134 TGGCCAAGCACAAAGCCATAGAGAAAAGGGAGAGCCGTGAGCCAGGAAAGCCGAACTGCGAC 2193

Db 2523 TTGCTAAGTACAAAGTTAGCAGAAAAGACAGAGACC--AGCCATGAACAAATGGCTTTTCA 2579

QY 2194 ACAGGTTTTCGGCTGGAGGAGGCTAAAAAGTCTGTGATTTACAGGCGGAGGTGAGGCTCTCA 2253

Db 2580 AAAGGCTTCAAGAAGAAGAACTAAAGTCAAGGCACTCTCAAGAGAAGTGTGATGATTA 2639

QY 2254 AGGAGAAGATCCAGAGCTGATGAACAAGGAAGACAGCTGTCTCAGCTTCAAGTCGACT 2313

Db 2640 AAGAGAAAAATTCATGAATACATGGCAACTGAAGACCTAATATGTCACTCCAGGGAGATC 2699

QY 2314 ATTCGGTCTTCAGCAAAAGATTTATGG-AAGAAGAAAATAAGAAACAAGAAACATGGGGAGG 2372

Db 2700 ACTCAGTCTCGCAAAAAAAACTAAATCAACAAGAAAAACAGGAACAGAGATTTAGGAAGA 2759

QY 2373 GAGGTCTCAATCTGACCAAGGAGCTAGAGCTTTCCAAAGCGCTACAGCCGAGCTCTCAGG 2432

Db 2760 GAGATTGAAAACCTCACTAAGGAGTTAGAGAGGTACCGGCATTTTCAGTAAGAGCCTCAGG 2819

QY 2433 CCGAGTGGGAACCGCCGAAGGATGGTGGACGCTGTGGCCTCCACTGGGGTGCAGACC 2492

Db 2820 CCTAGTCTCAATGGAGAAGAAATTTCCGATCCTCAAGTATTTTCTAAAGAAAGTTCAGACA 2879

QY 2493 GAGCGGTGTGCGGGGATGCTGCGGAGGAGGAGACCCCGGCTGTGTTTCAATTCGCAATCC 2552

Db 2880 GAAGCAGTAGACAAATGAACCACTGATTAACAAGAGCCTCATTTCTCTGGAACTGCAGTC 2939

QY 2553 TTCCAGGAGGAAAATCACATCATGAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCATG 2612

Db 2940 ATCAATGGTCAGTTATATGAGGAGAGTGAGAATCAAGACGAGGACCCCTAATGA-----T 2993

QY 2613 GAACGGTCTCGTCTCGACAGGTATCCCCAGCAGCAGGAATGAGCTACCATGAGGAAG 2672

Db 2994 GAGGATCTGTGCTGCTCTCAATTCAGGAGGAGTCTACTTCCATGTCTGTTAAACAGAAAG 3053

QY 2673 TCTTGGATTCTTGGATGAGAAAAAGAGAAAACGGTCTTCCACTCCGCAGGAGAAAAGG 2732

Db 3054 CTATGGATTCCCTGGATGAAATCCAAGGAGGGCCCATCTTCAGAAATGGAATAATGCAA--- 3110

QY 2733 CCCAGGCCAAACAGGGTGCAGGGCACCCCGGGAGCTGGTCTCTAGCACCAAGCAGGGC 2792

Db 3111 ACTAAACCCCAATGCCAACTTTGTGCAACCTTGAGATCTAGTCTTAAGCCACACACCTGGG 3170

QY 2793 CAGCCCTTACATCCGTGTGACACCAGATCATGAGAAACAGCACTGCCACCTGGAGATC 2852

Db 3171 CAGCCACTTCATATAAAGGTTACTCCAGACCATGTACAAAACACAGCCACTCTTGAATC 3230

QY 2853 ACAAGCCCCACATCTGAAG-----AGTTTTTCTTAGTACCACCGTCAATTCCTACCTTA 2906

Db 3231 ACAAGTCCAACCAAGAGAGTCCCTCACTCTTACACGAGTACTGCAGTGATACCGAACTGT 3290

QY 2907 GGCAACCCAGAAAACCAAGAAATAACCATTTATTCATCACCCTCAATGTTCATGTCGCAAAAGCCC 2966

Db 3291 GGCACGCCAAAGCAAAAGGATAACCATCCTCCAAAACGCCTCCATAACACCCAGTAAAGTCC 3350

AAH06432;
26-JUN-2001 (first entry)
Human cDNA clone (5'-primer) SEQ ID NO:3267.
Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
Homo sapiens.
EP1074617-A2.
07-FEB-2001.
28-JUL-2000; 2000EP-00116126.
29-JUL-1999; 99JP-00248036.
27-AUG-1999; 99JP-00300253.
11-JAN-2000; 2000JP-00118776.
02-MAY-2000; 2000JP-00183767.
09-JUN-2000; 2000JP-00241899.
(HELI-) HELIX RES INST.
Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
WPI; 2001-318749/34.
Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
Claim 1; SEQ ID NO 3267; 2537pp + Sequence Listing; English.
The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
Sequence 844 BP; 328 A; 130 C; 211 G; 170 T; 0 U; 5 Other;
Query Match 13.3%; Score 581.6; DB 4; Length 844;
Best Local Similarity 81.1%; Pred. No. 4.4e-126;
Matches 685; Conservative 0; Mismatches 159; Indels 1; Gaps 1;
QY 1719 GAAAGCAAGAGCTTTTAAAACTCAAACTGAAATGGAGGAAAGGAGTACAGTCTGACA 1778
Db 1 GNAAGTAAGAAACTTTTAAAACTAAAATCTGAAATGGAGGAAAGGATATACAACCTTGACA 60
QY 1779 AAGGAGAGGGATGAGCTGGGTAAACTGAGGAGCGAAGAAAGGTCCTGTGAACTG 1838
Db 61 AGAGAAAGAGATGAGTTGATAGGCAATTGAAAAGTGAAGAGAGAAAATCCTCTGAATTA 120

QY 1839 AGCTGCAGTGTAGACTTACTAAAGAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGGGAA 1898
Db 121 AGCTGCAGTGTGACTTACTAAAGAAGAGACTTGATGGTATAGAGGAAGTGAAAGAGAA 180
QY 1899 ATAAACCGAGGTAGTCTGTCGAAGGGGTCTGAGTTCACTCTGCCCGAAGACAATAAGATC 1958
Db 181 ATAACAAGAGGAAGGTCACGAAAAGGCTTGAGCTCAGCTGCCCGAAGATAATAAGATT 240
QY 1959 AGAGAACTAACGCTTGAAATCGAGAGACTGAAGAAAACGGCTCCAGCAGTTGGAGGTGGTG 2018
Db 241 AAGGAACTAACACTTGAAATTGAGAGACTGAAGAAAACGTCTCCAACAATTGGAAAGTGTC 300
QY 2019 GAGGGGACTTGTATGAAGACCGGAGGACGAATATGACAGATTGGAGCAGAAAGTTTCAGAAC 2078
Db 301 GAAGGGGATTGTATGAAGACAGAAAGATGAGTATGATCAGCTGGAACAGAAAATTTAGAACT 360
QY 2079 GAGCAGGATAAGGCAAACTTCTCTCCAGCAGCTCGAGGAAAATCAAACACCAAATGGCC 2138
Db 361 GAGCAGGATAAGGCTAACTTCTCTCTCAACAACTAGAGGAGATCAAGACACCAAATTGCC 420
QY 2139 AAGCAAAAGCCATAGAGAAAAGGGAGGCCGTGTAGCCAGGAAGCCGAACTGCGACACAGG 2198
Db 421 AAGATAAAGCAATAGAGAAAGGTGAGGTTGTGAGCCAGGAAGCTGAACTGAGACACAGA 480
QY 2199 TTTTCGCTGGAGGAGGCTAAAAGTCGTGATTTTACAGCCGAGGTGCAGGCTCTCAAGGAG 2258
Db 481 TTTTCGCTTGGAAAGAGCTAAAAGTCGAGACTTAAAAGCCGAAGTACAAGCTCTTAAAGAG 540
QY 2259 AAGATCCACGAGCTGATGAACAAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTATTTCG 2318
Db 541 AAGATTACGAATTAAATGAACAAAGAAAGATCAGCTTTCTCAGCTCCAGGTAGATTATCT 600
QY 2319 GTCCTTCAGCAAAAGATTATGGAAGAAGAAAACCTAAGAACAAAGACATGGGGAGGAGGTC 2378
Db 601 GTACTTCAACAAAGATTATGGAAGAAGAAAATAGAACAAAACATGGGGCANGAGGTT 660
QY 2379 CTCATCTGACCAAGAGCTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGT 2438
Db 661 CTCATCTGACCAAGAGTTGGAGCTTTCCAAGCGCTACAGCAGAGCTCTTA-GCCCACT 719
QY 2439 GGGAAAGGCCGAGGATGGTGGACGTGCCTGTGGCTCCACTGGGGTGCAGACCAGGCG 2498
Db 720 GTGAATGGAAGAAGAAATGGTGGATGTTCTCTGTGACGTCAACTGGAGTCCAAACTGATGCA 779
QY 2499 GTCTCGGGGATGCTGCGGAGGAGGAGAGACCCCGGCTGTGTTCAATTCGAAAATCCTTCCAG 2558
Db 780 GTCCAGCGGTGAANCAACAGANGAAGAAAACGCCAGCTGATTTCATACGGAATCCTTNCAG 839
QY 2559 GAGGA 2563
Db 840 GAAGA 844
RESULT 15
AAI60958
ID AAI60958 standard; cDNA; 1326 BP.
XX
AC AAI60958;
XX
DT 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 4947.
DE Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.

THIS PAGE BLANK (003710)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2004, 06:32:00 ; Search time 3716 Seconds
(without alignments)
5837.425 Million cell updates/sec

Title: US-10-788-793-1
Perfect score: 4364
Sequence: 1 ccactgggtttcttcaggga.....aaaaaaaaaaaaaaaaaa 4364

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3267054 seqs, 2485319735 residues

Total number of hits satisfying chosen parameters: 6534108

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10:	/cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15:	/cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16:	/cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18:	/cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
19:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3025.6	69.3	4892	15	US-10-037-270-1065	Sequence 1065, Ap
2	3025.6	69.3	4892	16	US-10-117-722-1065	Sequence 1065, Ap
3	1527.6	35.0	2402	13	US-10-116-802-180	Sequence 180, App
4	1191.2	27.3	1798	16	US-10-108-260A-1637	Sequence 1637, Ap
5	790.2	18.1	3415	15	US-10-309-851-13	Sequence 13, Appl
6	788.6	18.1	3416	15	US-10-309-851-15	Sequence 15, Appl
7	788.6	18.1	3430	15	US-10-309-851-11	Sequence 11, Appl
8	777.6	17.8	3998	15	US-10-309-851-9	Sequence 9, Appl
9	565	12.9	3069	16	US-10-108-260A-1126	Sequence 1126, Ap
10	559.2	12.8	3587	14	US-10-071-766-132	Sequence 132, App
11	541.6	12.4	3025	10	US-09-960-706-972	Sequence 972, App
12	541.6	12.4	3025	10	US-09-873-319-634	Sequence 634, App
13	541.6	12.4	3025	15	US-10-171-311-47	Sequence 47, Appl
14	496.6	11.4	2355	15	US-10-309-851-23	Sequence 23, Appl

15	314.4	7.2	1258	10	US-09-814-353-19975	Sequence 19975, A
16	292	6.7	1050	15	US-10-309-851-7	Sequence 7, Appl
17	265.4	6.1	795	15	US-10-309-851-5	Sequence 5, Appl
18	248.2	5.7	1568	9	US-09-925-302-52	Sequence 52, Appl
19	248.2	5.7	1568	13	US-09-925-302-52	Sequence 52, Appl
20	240.4	5.5	1503	15	US-10-198-846-13111	Sequence 13111, A
21	239.8	5.5	1625	16	US-10-104-047-1383	Sequence 1383, Ap
22	204.4	4.7	720	15	US-10-309-851-35	Sequence 35, Appl
23	202.8	4.6	720	15	US-10-309-851-3	Sequence 3, Appl
24	182.8	4.2	426	9	US-09-864-761-18880	Sequence 18880, A
25	154	3.5	1152	15	US-10-309-851-37	Sequence 37, Appl
26	154	3.5	1158	15	US-10-309-851-21	Sequence 21, Appl
27	153	3.5	461	10	US-09-918-995-11837	Sequence 11837, A
28	130	3.0	423	9	US-09-962-436-477	Sequence 477, App
29	126.6	2.9	497	9	US-09-864-761-2138	Sequence 2138, Ap
30	115.2	2.6	542	10	US-09-814-353-14401	Sequence 14401, A
31	109.2	2.5	400	15	US-10-198-846-6071	Sequence 6071, Ap
32	109.2	2.5	850	15	US-10-198-846-6373	Sequence 6373, Ap
33	93.6	2.1	2584	16	US-10-120-988-389	Sequence 389, App
34	93.6	2.1	3489	13	US-09-894-273-1	Sequence 1, Appl
35	93.6	2.1	3489	15	US-10-294-804-1	Sequence 1, Appl
36	92.6	2.1	216	13	US-10-085-783A-25432	Sequence 25432, A
37	92.6	2.1	216	16	US-10-242-535A-25432	Sequence 25432, A
38	86.6	2.0	463	10	US-09-814-353-1661	Sequence 1661, Ap
39	86.6	2.0	463	10	US-09-814-353-8017	Sequence 8017, Ap
40	86.4	2.0	4818	16	US-10-367-687-15	Sequence 15, Appl
41	82	1.9	2479	16	US-10-108-260A-1966	Sequence 1966, Ap
42	80.2	1.8	2093	16	US-10-104-047-1666	Sequence 1666, Ap
43	72.2	1.7	2480	16	US-10-104-047-576	Sequence 576, App
44	72.2	1.7	5397	13	US-10-112-944-128	Sequence 128, App
45	70.6	1.6	1799	16	US-10-108-260A-1885	Sequence 1885, Ap

ALIGNMENTS

RESULT 1
US-10-037-270-1065
; Sequence 1065, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillingshast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: NO. US20030104529A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt FL_genes Version 1.0
; SEQ ID NO 1065
; LENGTH: 4892
; TYPE: DNA
; ORGANISM: Homo sapiens

; FEATURE:									
; NAME/KEY: CDS									
; LOCATION: (634)..(4275)									
US-10-037-270-1065									
Query Match									
Best Local Similarity 69.3%; Score 3025.6; DB 15; Length 4892;									
Matches 3593; Conservative 0; Mismatches 734; Indels 27; Gaps 10;									
Qy	1	CCACTGGGTCTCTCAAGGGATAAACCGCGGC-GAAAGAACACACCATTGGTT-AAGGAG	58						
Db	558	CCACTGGGTCTCTCAAGGGATAAACTACCTACATAGAGGACATACCCTTTGGTTAAAGGAG	617						
Qy	59	TCGACAAACAGGTGGGAATGAGATCACGAAATCAAGGTGGAGAAAAGTTTCATCTAAACGGGCA	118						
Db	618	CTGCCGTGAGGTGGGAATGAGATCTCGAAACCAAGGTGGTGAAGTGCATCTGATGGGCA	677						
Qy	119	TGTCCTCGCCCCAAGTCTCCATCATCAGCAGTGATGGTGGTAAGGGCCCTCAGAAGA	178						
Db	678	TATCTCTGTCCCAAGCCCTCCATCATCGGCAATGCTGGTGAAAAAAGTCTCTCAGAAGA	737						
Qy	179	TGCAAAAAAGAAACAAG---GCCAATCGGAAGGAGGAGGATGTCTATGGCTTCGGGAACATAT	235						
Db	738	TGCAAAAAAGAAAGAAATCAAAATAGGAAGGAGGATGTATGTCATGGCCTCAGGAACCTGT	797						
Qy	236	CAAAAGGCACCTCAACCATCTGGAGAAAGTGAGAAAAAGACTAAGAAAGTCTGTGGAGTT	295						
Db	798	CAACGACACCTTAAAAACATCTGGAGAAATGTGAACGAAAAAACAATAAGAAATCCCTGGAGTT	857						
Qy	296	ATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGGAGTTGCAGGCTCGAGA	355						
Db	858	ATCCAAAGAAGACCTCATCCAACTACTCTAGTATAATGGGAAGGGAGTTGCAGGCCCAGAGA	917						
Qy	356	AGATGTCTCCACATGCTGAGGACAGAGAAAAACCAAGCCCGAGGTTCTCGAGGGCACACTA	415						
Db	918	AGATGTGATCCACATGCTGAAGACACAGAGAAAAACCAAGCCTGAGGTTCTGGAGGCTCATTA	977						
Qy	416	TGGATCTGCAGAAACCTGAGAAAGTGCTTCGGGTCTTGACCGGAGATGCCATCCTTGCTCA	475						
Db	978	CGGGTCTCGGAGCCAGAGAAAGTGCTGCGGGTCTCTGCACCGAGATGCCATCTTTGCCCA	1037						
Qy	476	AGAGAAGTCCATAGGAGAAAGACGTCTATAGAAACCTATCTCAGAGCTTGACAGACTGGA	535						
Db	1038	GGAGAAATCCATAGGAGAAAGATGTCTATGAGAAACCGATTTCAGAGCTGGACAGACTTGA	1097						
Qy	536	GGAAAAAGCAGAAAGGAGACGTACCGCCGCATGCTAGAGCAGCTGCTGGCTGAGAAAGTG	595						
Db	1098	GGAAAAACAGAAAGAAACCTACCGGGCGCATGCTAGAGCAGCTGTTGCTGGCCGAGAAAGTG	1157						
Qy	596	TCACAGGGCCACCGTGTACGAGCTGGAGAACGAGAGAACCAAGCACACACTGACTACATGAA	655						
Db	1158	TCATAGGCGCACCGTATACGAGTTAGAGAACGAGAGAGCATAAACACACTGACTACATGAA	1217						
Qy	656	CAAGAGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGAGAGAGGTTTGAAAAAAGCTCCT	715						
Db	1218	CAAGAGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGGAGAGGTTTAAAAAAGCTCCT	1277						
Qy	716	TGAACAAGAAAAAGCTTACCAAGCCCGCAAGAAAGAAACGCTAAGCGGCTCAACAA	775						
Db	1278	TGAACAAGAAAAAGGCTTATCAAGCCCGCAAGAAAGAAAGGAAATGCTAAACGACTCAATAA	1337						
Qy	776	ACTTCGAGATGAGCTTGTGAAGCTCAAGTCCCTTCGCCCTCATGTTGGTGGACGAGAGGCA	835						
Db	1338	ACTAAGAGATGAGCTTGTTAAACTCAAAATCCTTTGCACTCATGCTGGTGGATGAAAGACA	1397						
Qy	836	GATGCACATCGAGCAACTGGGCGCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAAGCT	895						
Db	1398	AATGCACATTGAACAACCTTGGCCCTGCAAGCCAGAAAGTACAGGATCTTACTCAGAAAGCT	1457						
Qy	896	GAGGGAGGAGAAAGAAAAACTCAAAGCGGTCACTTACAAATCCAAGGAAGACCGCCAGAA	955						
Db	1458	GAGGGAAGAAAGAGAGAGCTCAAAGCCCATTACTTCCTCAAATCCAAGAAAGACAGACAGAA	1517						

Qy	956	GCTGCTCAAGTTAGAAGTGGACTTCGAACAACAAGGCCTCGAGGTTTTTCCAGGAGCACGA	1015						
Db	1518	ATTGCTCAAGTTAGAAGTGGACTTTTGAACAACAAGGCTTCGAGGTTTTTCTCAAGAGCATGA	1577						
Qy	1016	AGAGATGAACCGCCAAATTTGGCGAATCAAGAATCTCACACCGGCAACTTCGACTCAAACT	1075						
Db	1578	AGAGATGAACCGCTAAACTGGCTAATCAAGAGTCTCACATATAGGCAACTTAGACTCAAGCT	1637						
Qy	1076	GGTTGGCTTATCGAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAAAGCCTTCAGAAGGC	1135						
Db	1638	GGTTGGCTTAAACCCAAAGAATCGAGGAGCTAGAAGAGACCAACAAAAATCTGCAGAAGGC	1697						
Qy	1136	AGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAACCTCCAG	1195						
Db	1698	AGAGGAAGAACTTCAAGAATTAAAGAGATAAAATTGCCAAAGGAGAAATGTGGAACCTCTAG	1757						
Qy	1196	TCTCATGGCGGAAGTGGAGAGTCTGCCAAAGCGCGTGTCTTGAGATGGAGGGCAAGGATGA	1255						
Db	1758	CCTCATGGCAGAAGTGGAAAAATCTTCGAAAAGCGTGTGCTTGAATTGAAGGTAAAGATGA	1817						
Qy	1256	AGAGATCACGAAGACCGAGGCCAGTGC CGGGAGCTGAAGAAGAAAGCTCCAAGAGGAAGA	1315						
Db	1818	GGAGATCACTAAAACTGAATCCCAGTGTAGGGAAATTGAGGAAGAAAGCTGCAAGAGGAAGA	1877						
Qy	1316	ACACCCACAGCAAGGAACCTTAGACTAGRAAGTGGAGAAAGTGCAGAAAGAGGATGTCTGAGCT	1375						
Db	1878	ACACCATAGTAAGGAGCTCAGACTTGAAGTTGAGAAAGCTACAGAAGAGAAATGTCTGAACCT	1937						
Qy	1376	GGAGAACTGGAGGAAGCGTTTCAGCCGAGTAAGTCGGAATGCAACCCAGCTCCATCTGAA	1435						
Db	1938	AGAGAAATTGAAGAAAGCATTTAGCAAGAGTAAATCTGAGTGCACCCAGCTACATTTAA	1997						
Qy	1436	CCTGGAGAAAGGAGAGAAACCTTAACCAAGA CCTGCTGAACGAGCTGGAGGTGGTCAAGAG	1495						
Db	1998	TCTGGAGAAAGAAAGAACTTAAACCAAGACCTGCTAAATGAATTGGAGGTGGTCAAGAG	2057						
Qy	1496	TCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTCGAGAAGGCCGAGTTAAGCCTCAA	1555						
Db	2058	TCGAGTTAAAGAAATTGGAATGTTCTGAAAGTAGATTGGAAAAAGGCTGAATTAAGCCTAA	2117						
Qy	1556	AGATGACCTTACAAAGCTGAAAGTCCTTCACTGTGATGCTGGTGGATGAGAGGAAAAATAT	1615						
Db	2118	AGATGATCTTACCAAGTTGAAAGTCATTTACCGTGTGCTGGTGTGATGAAGGAAAAATAT	2177						
Qy	1616	GATGARGAAAAATAAGCAAGAAGAGAGGAAAGTGGAATGGGTTGAATAAAAACTTTAAGGT	1675						
Db	2178	GATGGAATAAAATAAAACAAGAAGAGAGAAAAAGTGATGGACTCAA TAAAAATTTTAAGGT	2237						
Qy	1676	GGAGCAGGGAAAAAGTCATGGATGTGACGGAAAAAGCTAATCGAGGAAAGCAAGAAAGCTTTT	1735						
Db	2238	GGAACAAGGAAAAAGTTATGGATGTAAC TGAANAAC TAATTGAAGAAAGTAAGAAACCTTTT	2297						
Qy	1736	AAACTCAAAATCTGAAATGGAGGAAAGGAGTACAGTCTGACAAAAGGAGGGATGAGCT	1795						
Db	2298	AAACTTAAATCTGAAATGGAGGAAAAAGTATACAAC TTAGCAAGAGAAAGAGATGAGTT	2357						
Qy	1796	GATGGTTAAACTGAGGAGCGAAGAAGAAAAGGTCC TGTGAAC TGAGCTGCAGTGTAGACTT	1855						
Db	2358	GATAGGCAAAATTGAAAAGTGAAGAAGAAAAATCC TCTGAATTAAGCTGCAGTGTGACTT	2417						
Qy	1856	ACTAAAGAAAGCGGCTTGATGGCATAGAGGAGGTAGAAA GGGAAATAAACCGAGGTAGGTC	1915						
Db	2418	ACTAAAGAAAGAGACTTGATGGTATAGAGGAAGTGAAGAAGAGAAATAACAAGAGGAAGGTC	2477						
Qy	1916	GTGCAAGGGGTCTGAGTTCACTGCCCGGAAGACATAA GATCAGAGAACTAACCGCTTGA	1975						
Db	2478	ACGAAAAGGGTCTGAGCTCACCTGCCCGGAAGATAA TAAGATTAAAGGAAC TAACACTTGA	2537						
Qy	1976	AATCGAGAGACTGAAGAAAAGCGCTCCAGCAGTTGGAGTGGTGGAGGGGACTTTGATGAA	2035						
Db	2538	AATTGAGAGACTGAAGAAAAGCTCTCCAACAATTGGAAGTGGT CGAAGGGGATTTGATGAA	2597						
Qy	2036	GACCGAGGACGAATATGACCAGTTGGAGCAGAAAGTT CAGAAACCGAGCAGGATAAGGCCAA	2095						

Db 2598 GACAGAAAGATGAGTATGATCAGCTGGAACAGAAATTTAGAACTGAGCAGGATAAAGGCTAA 2657
QY 2096 CTTCTCTCTCCAGCAGCTCGAGGAAATCAAACACCAAATGGCCAAAGCACAAAGCCATAGA 2155
Db 2658 CTTCTCTCTCAACAACACTAGAGGAGATCAAGCACCAAAATGGCCAAAGAATAAAGCAATAGA 2717
QY 2156 GAAAGGGGAGGCCGTGAGCCAGGAAGCCGAACTGCGACACAGGTTTTCGGCTGGAGGAGGC 2215
Db 2718 GAAGGTGAGGTTGTGAGCCAGGAAGCTGAACTGAGACACAGATTTCCGGTTGGAAGAAGC 2777
QY 2216 TAAAAGTCGTGATTTACAGCCCGAGGTGCAGGCTCTCAAGGAGAAGATCCACGAGCTGAT 2275
Db 2778 TAAAAGTCGAGACTTAAAAGCCGAAAGTACAAGCTCTTAAAGAGAAGATTCACGAATTAAT 2837
QY 2276 GAAACAAGGAAGACCAAGCTGTCTCAGCTCCAAGTCGACTATTCCGTCTTCAGCAAAAGATT 2335
Db 2838 GAACAAAAGNAGATCAGCTTTCTCAGCTCCAGGTAGATTATTCTGTACTTCAACAAAAGATT 2897
QY 2336 TATGGAAGAAGAAACTAAGAACAAGAACATGGGGAGGGAGGTCTCTCAATCTGACCAAGGA 2395
Db 2898 TATGGAAGAAGAAATAAGAACAATAAATCATGGGGCAGGAGGTTCTCAATCTGACCAAGA 2957
QY 2396 GCTAGAGCTTTTCCAAAGCGCTACAGCCGAGCTCTCAGCCGAGTGGGAACGGCCGAAGGAT 2455
Db 2958 GTTGGAGCTTTTCCAAAGCGCTACAGCAGAGCTCTTAGGCCAGTGTGAATGGAAGAAGAT 3017
QY 2456 GGTGGAGCTGCTGTGGCTCCACTGGGCTGAGACCGAGGCGGTGTGCGGGGATGCTGC 2515
Db 3018 GGTGGATGTTCTGTGACGTCAACTGGAGTCCAACTGATGCAGTCAGCGGTGAAGCAGC 3077
QY 2516 GGAGGAGGAGACCCCGCTGTGTTTATTTCGCAATCTCTCCAGGAGGAAATCACATCAT 2575
Db 3078 AGAGGAAGAAACGCCAGCTGTATTATACGGAATCTCTCAGGAAGAAATCATATTAT 3137
QY 2576 GAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCATGGAACGGTCTCGGTCTCGACAG 2635
Db 3138 GAGTAATCTTCGGCAGGTGGGATTGAAGAAACCCCGTGAAGAAGATCTTCTGTCTAGACAG 3197
QY 2636 GTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTGGATTCTCTTGGATGAGAAA 2695
Db 3198 GTATCTCCAGCAGCAAAATGAGCTCACTATGAGAAAGTCTTGGATTCCATGGATGAGAAA 3257
QY 2696 AAGAGAAAACGGTCTTCCACTCCGACGAGAAAGGGCCAGGCCAAACCAAGGTGCAGG 2755
Db 3258 GAGGGAACACGGCCCTCCATCACTCAGGAGAAAGGGCCCCCGAACAAATTCAGTCCAGG 3317
QY 2756 GCACCCCGGAGCTGGTCTTAGCACCNAAGCAGGGCAGCCCCCTACACATCCGTGTGAC 2815
Db 3318 GCACCCAGGAGAGGTAGTCTTTCACCAAGCAGGGCCAGCCCCCTGCATATTCCAGTGAC 3377
QY 2816 ACCAGATCATGAGAACAGCACTGCCACCTGGAGATCACAAGCCCCCACATCTGAAGAGTT 2875
Db 3378 ACCAGACCAGAAACAGCACTGCGACTTTGGAGATAACAAGCCCCGACATCTGAAGAATT 3437
QY 2876 TTTCTCTAGTACCACCGTCACTTCCCTTAGGCAACCAAGAACCAAGAATAACCATTTAT 2935
Db 3438 TTTTCTAGTACCACCTGTCTTCTTACCTTAGGGAAATCAGAAACCAAGAATAACCATTTAT 3497
QY 2936 TCCATCACCCCAATGTCTATGTCGCAAAAGCCAAAGTCAGATCTCTCTCGGCCCAGA 2995
Db 3498 TCCATCACCAACCGTTATGCTCTCAAAAACAAAAAGTGGAGATACTACTCTTGGCCCCAGA 3557
QY 2996 ACAGGCCATGTCCCTGTACAGATTACTACTATTTCCAGAGAGAAGAGCCCGGAAGGTGG 3055
Db 3558 ACAGGCCATGTCCCCAGTCACAAATTACTACATTTTCAGAGAGAAGACTCCAGAAAAGTGG 3617
QY 3056 AAGGAGCGCCTTTGCGACAGGCCCTGCATCCCCCATCCAAATCATGACGGTGTCAACATC 3115
Db 3618 AAGAGCGCATTTGACAGACAGGCCCAATCCCTTATTCAGATAATGACGGTGTCTACATC 3677
QY 3116 TGCAGCTCCCACCTGAAATCGTGTCTCTCTGAAATCTCAGGAAGTGCCTATGGGAAGGAC 3175

Db 3678 AGCAGCACCACTGAGATTGCAGTTTCTCCCGAATCCCAGGAAATGCCCATGGGACGGAC 3737
QY 3176 TATCCTCAAAGTCACCCCGGAAAAACAACAACTGTTCCAGCCCCCGTCCGGAAGTACAACTC 3235
Db 3738 AATCCTCAAAGTCACCCCAGAAAAACAGACTGTTCCAACTCCAGTACGGAATAACAACTC 3797
QY 3236 CAATGCTAATATCATCACACGGAAGACATAAATAATTCAATTACCTGGGTTCTCAGTT 3295
Db 3798 CAATGCCAATATATAAACACACAGAGGACAATAAATAATTCAATTACCTTAGGGTCTCAGTT 3857
QY 3296 TAAGCGATCTCCTGGGCCTGCCGCTGAAGGCGTGAGCCAGTTATACCGTCCCGCCTGT 3355
Db 3858 TAAACGGTCCCCTGGGACTTCAGGTGAAGGAGTCAGTCCAGTTATTACTGTCCGACCAGT 3917
QY 3356 CAACGTGACAGCGGAGAAGGAGGTTTCTACAGGCACAGTCCCTTCGCTCTCCAGGAACCA 3415
Db 3918 AAACGTGACAGCGCAAAAGGAGTTTCCACCGGCACTGTCTTCGCTCTCCAGGAATCA 3977
QY 3416 CCTCTCTTCAAGACCCGGTGCTAGCAAAAGTGACCAAGCACTATAACTATAACCCCGGTCA 3475
Db 3978 CCTCTCTCACGGCCTGGTGCAAGCAAAAGTGACGACACTATCACCATAAACCCGGTCAC 4037
QY 3476 AACGTATCCACAGAGGAACCCAAATCAGTGTGAGGACAAGATGGGTCTATCTCAGCGGCC 3535
Db 4038 AACGTATCTGCTCGAGGAACCCAGTCAGTGTGAGGACAAGACGGGTCTATCCAGCGGCC 4097
QY 3536 TACCCCCACCCGATTCCTATGTCAAAAGGTATGAAAGCTGAAAGCCAGTAGTGGCAGC 3595
Db 4098 TACACCCACCCGATTCCTATGTCAAAAGGTATGAAAGCAGGAAGCCAGTAGTGGCAGC 4157
QY 3596 CTCAGGAGCAGGAAATCTGACCAAAATTCAGCCCTCGAGCTGAGACTCAGTCTATGAAAAT 3655
Db 4158 CCCAGGAGCAGGAAATCTGACCAAAATTCAGCCCTCGAGCTGAGACTCAGTCTATGAAAAT 4217
QY 3656 AGAGCTGAAGAAATCTGCAGCCAGCAGCACTGCCTCTCTTGGAGGGGGGAAGGGCTGAGG 3715
Db 4218 AGAGCTGAAGAAATCTGCAGCCAGCAGCACCACTCTCTCGAGGGGGGAAGGGCTGAGG 4277
QY 3716 GCAGTGGCTAAGGGGTATGTTGTAAGGATGCTACTGCTGCAGTGGAAACAAACCTTCTCT 3775
Db 4278 GCAGTGGCTAAGGGGTATGTTGTGAGATGCTACTGCTGCCGTGAAAGTGAAACCTTCAT 4337
QY 3776 CTGTGCCAACCCCTTCTCTTGTA - CTACTAAATTTAAAGTTTAAATATCTTGTATATAAAAT 3834
Db 4338 CTGTTTGTGCCAGTTCTTTACATGTACTAATTTAAAGTTTAAATATTTGTGTATATAAAAT 4397
QY 3835 AACCATTTAATAGCCATGCACCCCCCTCCCATTTTGTGATCTGTTCAATGACGGGAA 3894
Db 4398 AACCAACTAATAACCATTTGTCTT - TCCCATTTTGTGATTTGTTTGTGCTGGGAA 4455
QY 3895 TAGAATTAATTAGCAGAAATTTCTGTTTGTGAATGTTCTGTTGAAGATGTTGGTCCAGTT 3954
Db 4456 CA - - - - AAATTAGCAAAAACATAATTGCTTGTCTGCTAGAACCCAGGGGTGTTCTAGTTC 4511
QY 3955 CAGTTTACTTCTAGCATGTGGCCCCATTCAGAGGTAGCTCA CGAGTTGTGAAGCCCTCAA 4014
Db 4512 CAGTTTGTCTTCTAGCAAGTGGAACCCATCAATAGACCCCATCTGAG - - - - - CCTG 4560
QY 4015 TATCGTCACCGGAGAGATTGAGGACCACATPACATATATGCTCCCAAGGCTGGCTCCCAA 4074
Db 4561 TTTCTCTCATCAGTTAGATGTGGGACTCAAT - - CACACGCTCTTCAAGTCCGCTCCCAT 4618
QY 4075 TTTTCTCTAATTGPAAGCCAACTTTAATAGACTCAGTTCTGTGAT - TTTTCTTCTCCAAAA 4133
Db 4619 ATTTCTTAATTGCAAGCCAAATTTAATGTAACCTTGTTCACAAATAATTTTATTAAAAA 4678
QY 4134 AAAAATATTTTGAATAGGACAGAGTTTAAAGTTTGCATTTTGCATATCAAGCCCATGA 4193
Db 4679 AATCCTATTACAAAATAAGACATACTTTAACTATTGTCTTTGCTCTTTCACATCATGA 4738
QY 4194 GTTTGATATATGGGTTATAAGAAA - AGAATACTTTTACAGACTATTCACAGGCTCTCTAAAC 4252
Db 4739 ATTTGCTTTATGTGCTGGAAAAAACATCATATAGCTATTCACAGGCGCTGGACCTCTAAAA 4798

QY	4253	TTTTGGAACAAACAAAGCCCTTAATATGACCTCAGGAACAAATTTGAACATGAAATAAAA	4312
Db	4799	TTTTGCAACAAACAAAGGTTCTAAGATGATTTCAGGAATAAATGTGAACATGTAATAAAA	4858
QY	4313	TGGAATGAACCTGTGGAATCTTTAAAAA	4346
Db	4859	TGGAATGAATAATGGAATCCTAAAAA	4892
RESULT 2			
US-10-117-722-1065			
; Sequence 1065, Application US/10117722			
; Publication No. US20030219744A1			
; GENERAL INFORMATION:			
; APPLICANT: Tang, Y. Tom			
; APPLICANT: Liu, Chenghua			
; APPLICANT: Asundi, Vinod			
; APPLICANT: Zhang, Jie			
; APPLICANT: Drmanac, Radoje T.			
; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and			
; FILE REFERENCE: 784CIP2BCIP			
; CURRENT APPLICATION NUMBER: US/10/117,722			
; CURRENT FILING DATE: 2002-04-04			
; PRIOR APPLICATION NUMBER: 09/620,312			
; PRIOR FILING DATE: 2000-07-19			
; PRIOR APPLICATION NUMBER: 09/552,317			
; PRIOR FILING DATE: 2000-04-25			
; PRIOR APPLICATION NUMBER: 09/488,725			
; PRIOR FILING DATE: 2000-01-21			
; NUMBER OF SEQ ID NOS: 1104			
; SOFTWARE: pt_FL_genes Version 1.0			
; SEQ ID NO 1065			
; LENGTH: 4892			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (634)..(4275)			
US-10-117-722-1065			
Query Match 69.3%; Score 3025.6; DB 16; Length 4892;			
Best Local Similarity 82.5%; Pred. No. 0;			
Matches 3593; Conservative 0; Mismatches 734; Indels 27; Gaps 10;			
QY	1	CCACTGGGTTCTTCAAGGATAAAACGCGGC-GAAAGAACACACCATTGGTT-AAGGAG	58
Db	558	CCACTGGGTTCTTCAAGGATAAACTACCTACATAGAGCATACCTTTGGTTAAAGGAG	617
QY	59	TCGACAACAGGTGGGAATGAGATCACGAAATCAAGGTGGAGAAAGTTTCATCTAACGGGCA	118
Db	618	CTGCCGTCAGGTGGGAATGAGATCTCGAAACCAAGGTGGTGAAAGTGCACTGTATGGGCA	677
QY	119	TGTCTCCTGCCCCCAAGTCCTCCATCATCAGCAGTGATGGTGGTAAGGGCCCTCAGAAAGA	178
Db	678	TATCTCTGTCCCAAGCCCTCCATCATCGGCAATGCTGTTGAAAAAAGTCTCTCAGAAGA	737
QY	179	TGCAAAAAAGAACAAAG---GCCAATCGGAAGGAGGAGGATGTCTATGGCTTCGGAACTAT	235
Db	738	TGCAAAAAAGAAAGAAATCAAAATAGGAAGGAGGATGATGTCTATGGCCTCAGGAATGT	797
QY	236	CAAAAGGCACCTCAAAACCATCTGGAGAAAGTGAGAAAAAGACTAAGAACTCTGTGGAGTT	295
Db	798	CAAACGACACCTAAAAACATCTGGAGAAATGTGAACGAAAAAACAATAAGAAATCCCTGGAGTT	857
QY	296	ATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGGAGTTGCAGGCTCGAGA	355
Db	858	ATCCAAAGAAAGACCTCATCCAACTACTCAGTATAATGGAAGGGGAGTTGCAGGCCAGAGA	917
QY	356	AGATGTCATCCACATGCTGAGGACAGAGAAAAACCAAGCCCGAGGTTCTGGAGGCACACTA	415
Db	918	AGATGTGATCCACATGCTGAAGACAGAGAAAAACCAAGCCTGAGGTTCTTGAGGCTCATTA	977

QY	416	TGGATCTGCAGAAACCTTGAGAAAGTGTTCGGGTCTTCACCCGAGATGCCATCCTTGCTCA	475
Db	978	CGGGTCTGCGGAGCCAGAGAAAGTGTGCGGGTCTTCGACCGAGATGCCATCTTGGCCA	1037
QY	476	AGAGAGTCCATAGGAGAAAGACGTCTATGAGAAACCTTATCTCAGAGCTGGACAGACTGGA	535
Db	1038	GGAGAAATCCATAGGAGAAAGATGTCTATGAGAAACCGATTTTCAGAGCTGGACAGACTTGA	1097
QY	536	GGAAGACAGAAAGGAGACGTACCGCCGCATGCTAGAGCAGCTGCTGCTGGCTGAGAAAGTG	595
Db	1098	GGAATAACAGAAAGAAACCTACCGCGCATGCTAGAGCAGCTGTTGCTGCGCGAGAGTG	1157
QY	596	TCACAGGCGCACCGTGTACGAGCTGGAGAAACGAGAAAGCACAAAGCACACTGACTACATGAA	655
Db	1158	TCATAGCGCACCGTATACGAGTTAGAGAAACGAGAAAGCATAAACACACTGACTACATGAA	1217
QY	656	CAAGAGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGAGAGGTTGAAAAAGCTCCT	715
Db	1218	CAAGAGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGGAGAGGTTAAAAAAGCTCCT	1277
QY	716	TGAACAAGAAAAAGCTTACCAAGCCCGCAAAAGAAAAAGAAAAACGCTAAGCGGCTCAACAA	775
Db	1278	TGAACAAGAAAAAGGCTTATCAAGCCCGCAAAAGAAAAAGAAAAATGCTAAACGACTCAATAA	1337
QY	776	ACTTCAGATGAGCTTGTGAAGCTCAAGTCCTTCGCGCTCATGTTGGTGGACGAGAGGCA	835
Db	1338	ACTAAGAGATGAGCTTGTAAACTCAAATCCTTTTGACTCATGCTGGTGGATGAAAGACA	1397
QY	836	GATGCACATCGAGCAACTGGGCCTGCAGAGTCAAGAAAGTCCAGGACCTCAGTCTCAGAAGCT	895
Db	1398	AATGCACATTGAACAACTTGGCCTGCAAAAGCCAGAAAGTACAGGATCTTACTCAGAAGCT	1457
QY	896	GAGGAGGAGGAAGAAAAACTCAAAGCGGTCACTTACAAATCCAAGAGAGACCGCCAGAA	955
Db	1458	GAGGGAAGAAAGAGAGAGCTCAAAGCCATTACTTCCAAATCCAAAGAGACAGACAGAA	1517
QY	956	GCTGCTCAAGTTAGAAAGTGGACTTCGAACACAAAGGCTTCGAGGTTTTCCCAGGAGCACGA	1015
Db	1518	ATTGCTCAAGTTAGAAAGTGGACTTTGAACACAAAGGCTTCGAGGTTTTCTCAAGAGCATGA	1577
QY	1016	AGAGATGAACGCCAAATTGGCGAATCAAGAATCTCAAAACCGGCAACTTCGACTCAAACT	1075
Db	1578	AGAGATGAACCGCTAAACTGGCTAATCAAGAGTCTCAAAATAGGCAACTTAGACTCAAGCT	1637
QY	1076	GTTTGGCTTATCGAAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAGGC	1135
Db	1638	GTTTGGCTTAACCCAAAGAATCGAGGAGCTAGAAGAGACCAACAAAAATCTGCAGAAGGC	1697
QY	1136	AGAGAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAACCTCCAG	1195
Db	1698	AGAGGAAGAACTTCAAGAAATTAAGAGATAAAAATTGCCAAAGGAGAAATGTGGAAACTCTAG	1757
QY	1196	TCTCATGGCGGAAGTGGAGAGTCTGCCAAGCGCGTCTTGAGATGGAGGGCAAGGATGA	1255
Db	1758	CCTCATGGCAGAAGTGGAAAAATCTTCGAAAGCGTGTGCTTGAAATGGRAAGGTAAAGATGA	1817
QY	1256	AGAGATCACGAAGACCGAGGGCCAGTGCCTGGGAGCTGAAAGAGAAAGCTCCAAGAGGAAGA	1315
Db	1818	GGAGATCACTAAAACTGAATCCCAGTGTAGGGAATTGAGGAAGAAAGCTGCAAGAGGAAGA	1877
QY	1316	ACACCACAGCAAGGAACCTTAGACTAGAAGTGGAGAAAGCTGCAGAAAGAGGATGTCTGAGCT	1375
Db	1878	ACCCATAGTAAGGAGCTCAGACTTGAAGTTGAGAAAGCTACAGAAAGAGAAATGTCTGAACT	1937
QY	1376	GGAGAGCTGGAGGAAGCGTTTCAGCCCGAGTAAGTCGGAATGCACCCAGCTCCATCTGAA	1435
Db	1938	AGAGAAATTGAAGAAGCATTTAGCAAGAGTAAATCTGAGTGCACCCAGCTACATTTTAAA	1997
QY	1436	CCTGGAGAAGGAGAAAGAACTTAACCAAGACCTGCTGTAACCGAGCTGGAGGTGGTCAAGAG	1495
Db	1998	TCTGGAGAAAAGAAAGAACTTAACCAAGACCTGCTTAAATGAATTTGGAGGTGGTCAAGAG	2057

QY	1496	TCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAGGCCGAGTTAAGCCTCAA	1555
Db	2058	TCGAGTTAAAGAAATTGGAATGTTCTGAAAGTAGATTGGAAAAGGCTGAATTAAGCCTAAA	2117
QY	1556	AGATGACCTTACAAAAGCTGAAGTCCTTCACTGTGATGCTGGTGGATGAGAGGAAAAATAT	1615
Db	2118	AGATGATCTTACCAAGTTGAAGTCATTTACCGTGATGCTGGTTGATGAAAGGAAAAATAT	2177
QY	1616	GATGGAGAAAAATAAAGCAAGAGAGAGGAAAGTGATGGGTTGAATAAAAACTTTAAGGT	1675
Db	2178	GATGGAAAAAATAAACAAGAGAGAGAAAAAGTGATGGACTCAATAAAAAATTTAAGGT	2237
QY	1676	GGAGCAGGAAAAAGTCATGGATGTGACGGAAAAGCTAATCGAGGAAAGCAAGAGCTTTT	1735
Db	2238	GGAACAAGGAAAAAGTTATGATGTAACTGAAAAAACTAATTGAAGAAAGTAAGAACTTTT	2297
QY	1736	AAAACCTCAAATCTGAAATCGAGGAAAAAGGAGTACAGTCTGACAAAGGAGAGGGATGAGCT	1795
Db	2298	AAAACTAATAATCTGAAATCGAGGAAAAAGTATACAACTTGACAAGAGAAAGATGAGTT	2357
QY	1796	GATGGGTAAACTGAGGAGCGCAAGAAGAAAGGTCCTGTGAACTGAGCTGCAGTGTAGACTT	1855
Db	2358	GATAGGCATAATTGAAAAGTGAAGAAAGAAAAATCCCTCGAATTAAGCTGCAGTGTGACTT	2417
QY	1856	ACTAAAGAAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGGAAAAATAAACCGAGGTAGGTC	1915
Db	2418	ACTAAAGAAAGAGACTTGATGGTATAGAGGAAAGTGGAAGAGAGAAATAACAAGAGGAAGGTC	2477
QY	1916	GTCAAGGGGTCTGAGTTCACCTGCCGGAAGACAAATAAGATCAGAGAACTAACGCTTGA	1975
Db	2478	ACGAAAAGGGTCTGAGCTCACCTGCCGGAAGATAATAAGATTAAAGAACTAACACTTGA	2537
QY	1976	AATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTGGTGGAGGGGACTTGATGAA	2035
Db	2538	AATTGAGAGACTGAAGAAACGTCTCCAACAATTGGAAGTGGTCGAAGGGGATTTGATGAA	2597
QY	2036	GACCGAGGACGAATATGACCAGTTGGAGCAGAAGTTCAGAAACCGAGCAGGATAAGGCAAA	2095
Db	2598	GACAGAAGATGAGTATGATCAGCTGGAACAGAAATTTAGAACTGAGCAGGATAAGGCTAA	2657
QY	2096	CTTCCTCTCCCAGCAGCTCAGGAAATCAAACACCAATGGCCAAAGCACAAAGCCATAGA	2155
Db	2658	CTTCCTCTCTCAACAACCTAGAGGAGATCAAGCACCAAAATTGCCAAAGAATAAAGCAATAGA	2717
QY	2156	GAAAGGGGAGCGCTGAGCCAGGAAGCCGAACTGCGACACACAGGTTTCGGCTGGAGGAGGC	2215
Db	2718	GAAGGGTGAGGTTGTGAGCCAGGAAGCTGAACCTGAGACACAGATTTTCGGTTGGAAGAAGC	2777
QY	2216	TAAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAGGAGAAGATCCACGAGCTGAT	2275
Db	2778	TAAAAGTCGAGACTTAAAAGCCGAAGTACAAGCTCTTAAAGAGAAGATTTCAGCAATTAAT	2837
QY	2276	GAACAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTATTTCGGTCCTTCAGCAAGATT	2335
Db	2838	GAACAAGAAGATCAGCTTTCTCAGCTCCAGGTAGATATTCTGTACTTCAACAAGAGATT	2897
QY	2336	TATGGAAGAGAAACTAAGAACAAAGAACATGGGGAGGAGGTCTCTCAATCTGACCAAGGA	2395
Db	2898	TATGGAAGNAGAAAATAAGAACAAAAACATGGGCAGGAGGTTCTCAATCTGACCAAAAGA	2957
QY	2396	GCTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAAGGAT	2455
Db	2958	GTTGGAGCTTTCCAAGCGCTACAGCAGAGCTCTTAGGCCCAGTGTGAATGGGAAGAAGAT	3017
QY	2456	GGTGCAGCTGCTGCTCCACTGGGGTGCAGACCGAGCGGTTGTGCGGGGATGCTGC	2515
Db	3018	GGTGGATGTTCTCTGTGACGTCAACTGGAGTCCAAACTGATGCAGTCAGCGGTGAAGCAGC	3077
QY	2516	GGAGGAGGAGACCCCGGCTGTGTTTCATTTCGCAATCTCTCCAGGAGGAAAAATCACATCAT	2575
Db	3078	AGAGGAAGAAACGCCAGCTGTATTTCATACGGAAATCTCTTCCAGGAAGAAATCATATTAT	3137
QY	2576	GAGTAATCTTCGACAGGTAGGCCCTGAAGAAACCCATGGAACGGTCTCTCGGTCTCTCGACAG	2635

Db	3138	GAGTAATCTTCGGCAGGTGGGATTGAAGAAAACCCGTGGAAAGATCTTCTGTTCTAGACAG	3197
QY	2636	GTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTGGATTCTCTTGGATGAGAAA	2695
Db	3198	GTATCCTCCAGCAGCAATGAGCTCACTATGAGAAAGTCTTGGATTCCATGGATGAGAAA	3257
QY	2696	AAGAGAAAAACGGTCTCTCCACTCCGCAAGGAGAAAAGGGCCCCAGGCCAAACCCAGGGTGCAGG	2755
Db	3258	GAGGAAAAACGGCCCTCCATCACTCAGGAGAAAAGGGCCCCGGAACAAATTCAGTCCAGG	3317
QY	2756	GCACCCCGGGAGCTGGTCTTAGCACCAAAAGCAGGGCCAGCCCTACACATCCGTGTGAC	2815
Db	3318	GCACCCAGGAGAGGTAGTCTCTTTCACCAAAAGCAGGGCCAGCCCTGCATATTCGAGTGAC	3377
QY	2816	ACCAGATCATGAGAAACAGCACTGCCACCCCTGGAGATCAAAAGCCCCACATCTGAAGAGTT	2875
Db	3378	ACCAGACCACGAGAAACAGCACTGCGACTTTGGAGATAACAAGCCCCGACATCTGAAGAATT	3437
QY	2876	TTTCTCTAGTACCAACCGTCATTCCTACCTTAGGCAACCCAGAAACCAAGAAATAACCATTAT	2935
Db	3438	TTTTTCTAGTACCACTGTCTATCTCTACCTTAGGGAATCAGAAACCAAGAATAACCATTAT	3497
QY	2936	TCCATCACCCCAATGTCTATGTCGCAAAAAGCCCCAAAAGTGCAGATCCTACTCTCGCCCCAGA	2995
Db	3498	TCCATCACCAACGTTATGCCTCAAAAACAAAAAAGTGGAGATACTACTCTTGGCCCCAGA	3557
QY	2996	ACGAGCCATGTCCTGTACGATTACTACTATTTTCCAGAGAGAAGACCCCGGAAGGTGG	3055
Db	3558	ACGAGCCATGTCCTCAGTCACAATTACTACATTTTCCAGAGAGAAGACTCCAGAAAGTGG	3617
QY	3056	AAGGAGCGCCTTTGCCGACAGGCCCTGCATCCCCCATCCAAATCATGACGGTGTCAACATC	3115
Db	3618	AAGAGCGCATTTGCAGACAGGCCCCACATCCCTTATTCAGATAATGACGGTGTCTACATC	3677
QY	3116	TGCAGTCCCCACTGAAAATCGTGTCTCTCTCTGAATCTCAGGAAGTGCCCTATGGGAAGGAC	3175
Db	3678	AGCAGCACCACTGAGATTGCAGTTTCTCCCGAATCCAGGAAATGCCCATGGGACGGAC	3737
QY	3176	TATCTCAAAGTCACCCCGGAAAAACAACACTGTTCAGCCCCCGTCCGGGAAGTACAACTC	3235
Db	3738	AATCTCAAAGTCACCCCGAGAAAAACAGACTGTTCCAACTCCAGTACGGAATAACAACTC	3797
QY	3236	CAATGCTAATATCATCACCAAGGAGACAATAAAATTCACATTCACCTGGGTCTCAGTT	3295
Db	3798	CAATGCCAATATCATAAACCACAGAGGACAATAAAATTCACATTCACCTTAGGTCTCAGTT	3857
QY	3296	TAAGCGATCTCTGGGCCCTGCCGTGAAGCGGTAGCCAGTTATCACCGTCCGGCCTGT	3355
Db	3858	TAAACGGTCCCTGGGACTTCAGGTGAAGGAGTCAGTCCAGTTATTACTGTCCGACCAGT	3917
QY	3356	CAACGTGACAGCGGAGAAGGAGGTTTCTACAGGCACAGTCTCTCGCTCTCCAGGAACCA	3415
Db	3918	AAACGTGACAGCGCAAAAGGAGGTTTCCACCGGCACGTCTCTCGCTCTCCAGGAATCA	3977
QY	3416	CCTCTCTCAAGACCCCGTGTAGCAAAAGTGACCAAGTACCAAGCACTATAACTATACCCCGGTAC	3475
Db	3978	CCTCTCTCACGGCTGGTGCAAGCAAAAGTGACGAGCACTATCACCATAAACACCGGTAC	4037
QY	3476	AACGTCTCCACACAGGAACCCCAATCAGTGTGAGGACAAGATGGGTCTATCTCAGCGGCC	3535
Db	4038	AACGTCTATCTGCTCGAGGAACCCAGTCAGTGTTCAGGACAAGACGGGTCTATCCAGCGGCC	4097
QY	3536	TACCCCCACCCGATTCCTATGTCAAAAAGGTATGAAAGCTGGAAAGCCAGTAGTGGCAGC	3595
Db	4098	TACACCCACCCGATTCCTATGTCAAAAAGGTATGAAAGCAGGAAAGCCAGTAGTGGCAGC	4157
QY	3596	CTCAGGAGCAGGAAATCTGACCAAAATTCAGCCTCGAGCTGAGACTCAGTCTATGAAAAAT	3655
Db	4158	CCCAGGAGCAGGAAATCTGACCAAAATTCGAGCCTCGAGCTGAGACTCAGTCTATGAAAAAT	4217
QY	3656	AGAGCTGAAGAAATCTGCAGCCAGCACTGCCTCTCTTGGAGGGGGGAAGGCTGAGG	3715

Db 4218 AGAGCTGAAGAAATCTGCAGCCAGCAGCACCACTCTCTCGAGGGGGGAAGGGCTGAGG 4277
Qy 3716 GCAGTGGCTAAGGGGGTATGTTGTAAGGATGCTACTGCTGCAGTGGAAACAACCTTCTT 3775
Db 4278 GCAGTGGCTAAGGGGGTATGTTGTCAGATGCTACTGCTGCCGTGAAGTGAACTTTCAT 4337
Qy 3776 CTGTGCCAACCCCTTTCCTTGTA-CTACTAATTAAAGTTTAAATATCTTGTATTATAAAT 3834
Db 4338 CTGTTTGCCAGTTCTTTACATGTACTAATTTAAAGTTTAAATATTGTGTTTATAAAT 4397
Qy 3835 AACCATTTAATAGCCATGCACCCCTCCCATTTTGTCATCTGTTTCAATGCAGGGGAA 3894
Db 4398 AACCAACTAATAACCAATTGTCTT--TCCATTTTGTCATTTGTTTTCATGCTGGGGAA 4455
Qy 3895 TAGAATTAATTAGCAGAAATTCTGTTTGCTGAATGTTCTGTTGAAGATGTTGGTCCAGTT 3954
Db 4456 CA----AAATTAGCAAAACTATTGCTTGCTGCTTAGAAGCCAGGGCGTGTCTTAGTTC 4511
Qy 3955 CAGTTTFACTTCTAGCATGTGGCCCCCAATTCAAAGGTAGCTCACGAGTTGTGAAGCCCTCAA 4014
Db 4512 CAGTTTGTCTTAGCAAGTGGACCCCATCAATAGACCCATCTGAG-----CCTG 4560
Qy 4015 TATCGTACCGGAGAGATTTGAGGACCAATACATATATGCTCCCAAGGCTGGCTCCCAA 4074
Db 4561 TTTCTCATCAGTTAGATGTGGGACTCAAT--CACACGCTCTTCAAGTCCGGCTCCCAT 4618
Qy 4075 TTTTCCTAATTGTAAGCCAACTTTAATAGACTCAGTTCTGTGAT-TTTTTTTTCCAAAA 4133
Db 4619 ATTTCTTAATTGCAAGCCAAATTTAATGTACCTTGTTCCACAATAATTTTATTAAAAA 4678
Qy 4134 AAAAATATTTGAAATAGGACAGAGTTTAAACAGTTGTGCATTTTGCATATCAAGCCATGA 4193
Db 4679 AATCCTATTACAAAATAAGACATACTTTAACTATTGTCATTTGCCCTCTTTCACATCATGA 4738
Qy 4194 GTTTGATATATGGTTTATAAGAAA-AGRATACTTTTCAGAGCTATCACAGGGTCTCTAAAC 4252
Db 4739 ATTTGCTTATGTGCTGGAATAAACATCACATAGCTATCACAGGGCCTGGACCTCTAAAA 4798
Qy 4253 TTTTGGAAAAACAAAAGCCCTAATATGACCTCAGGAAACAATTTGAACATGAAATAAAA 4312
Db 4799 TTTTGCAAAAACNAAGGTTCTAAGATGATTTCAGGAATAAATGTGAACATGTAATAAAA 4858
Qy 4313 TGGAAATGAACACTGTGGAATCTTAAAAAATAAAAAA 4346
Db 4859 TGGAAATGAATAATGGAATCCTAAAAAATAAAAAA 4892

RESULT 3

US-10-116-802-180
; Sequence 180, Application US/10116802
; Publication No. US20030065157a1
; GENERAL INFORMATION:
; APPLICANT: Amy Lasek
; TITLE OF INVENTION: GENES EXPRESSED IN LUNG CANCER
; FILE REFERENCE: PA-0045 US
; CURRENT APPLICATION NUMBER: US/10/116,802
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 60/281,593
; PRIOR FILING DATE: 2001-04-04
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: PERL Program
; SEQ ID NO 180
; LENGTH: 2402
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 979448.2
US-10-116-802-180

Query Match 35.0%; Score 1527.6; DB 13; Length 2402;
Best Local Similarity 80.4%; Pred. No. 0;
Matches 1894; Conservative 0; Mismatches 439; Indels 23; Gaps 8;

Qy 1991 GAAACGGCTCCAGCAGTTGGAGGTGGTGGAGGGGACTTGATGAAGACCGAGGACGAATA 2050
Db 1 GAAACGCTCTCCAACAATTTGGAAGTGGTCTGAAGGGGATTTGATGAAGACAGAAAGATGAGTA 60
Qy 2051 TGACCAAGTTGGAGCAGAAATTTAGAACCCGAGCAGGATTAAGGCAAACTTCTCTCCCAGCA 2110
Db 61 TGATCAGCTGGAACAGAAATTTAGAACTGAGCAGGATTAAGGCTAACTTCTCTCTCAACA 120
Qy 2111 GCTCAGGAAATCAAACACCAAATGGCCAAAGCACAAAGCCATAGAGAAAGGGGAGGCCGT 2170
Db 121 ACTAGAGGAGATCAAGCACCAAATTTGCCAAGAATAAAGCAATAGAGAAGGTTGAGGTTGT 180
Qy 2171 GAGCCAGGAAGCCGAACTCGACACAGGTTTTCGGCTGGAGGAGGCTAAAAAGTCGTGATTT 2230
Db 181 GAGCCAGGAAGCTGAACTGAGACACAGATTTTCGGTTGGAAGAAAGCTAAAAAGTCGAGACTT 240
Qy 2231 ACAGCCGAGGTGCGAGCTCTCAAGGAGAAGATCCACGAGCTGATGAACAAGGAAGACCA 2290
Db 241 AAAAGCCGAAGTACAAGCTCTTAAAGAGAAGATTCACGAATTAATGAACAAGAAGATCA 300
Qy 2291 GCTGTCTCAGCTCCAAAGTCGACTATTTCGGTCTCTTCAGCAAAAGATTTATGGAAGAAGAAAC 2350
Db 301 GCTTTCTCAGCTCCAGGTAGATTAATCTGTACTTCAACAAAGATTTATGGAAGAAGAAAA 360
Qy 2351 TAAGAACAAAGACATGGGGAGGGAGTCTCAATCTGAACAAAGGAGCTAGAGCTTTCCAA 2410
Db 361 TAAGAACAAACATGGGGCAGGAGGTTCTCAATCTGAACAAAGAGTTGGAGCTTTCCAA 420
Qy 2411 GCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACCGSCGAAGGATGGTGGACGTCCTGT 2470
Db 421 GCGCTACAGCAGAGCTCTTAGGCCAGTGTGAATGGAAGAAGAATGGTGGATGTTCTCTGT 480
Qy 2471 GGCCTCCACTGGGGTGCAGACCGAGGCGGTGTGCGGGGATGCTGCGAGGAGGAGACCCC 2530
Db 481 GACGTCAACTGGAGTCCAAACTGATGCACTCAGCGGTGAAGCAGCAGAGGAAGAAACGCC 540
Qy 2531 GGCTGTGTTTCAATTCGCAAAATCCTTCCAGGAGGAAATCACATCATGAGTAATCTTCGACA 2590
Db 541 AGCTGTATTATACGGAATCCTTCCAGGAAGAAATCATATTATGAGTAATCTTTCGCA 600
Qy 2591 GGTAGCCTGAAGAAACCCCATGGAACGGTCTCGGTCTCTCGACAGGTATCCCCCAGCAGC 2650
Db 601 GGTGGGATGAAGAAACCCGTGGAAGATCTTCTGTCTAGACAGGTATCCTCCAGCAGC 660
Qy 2651 GAATGAGCTCACATGAGGAAGTCTTGGATTCCTTGGATGAGAAAAAGAGAAAAACGGTCC 2710
Db 661 AAATGAGCTCACTATGAGAAAGTCTTGGATTCATGATGAGAAAGAGGGGAAACCGGCC 720
Qy 2711 TTCCACTCCGAGGAGAAAGGGCCCGAGCCCAACCCAGGGTGCAGGGCACCCCGGGAGCT 2770
Db 721 CTCCATCACTCAGGAGAAAGGGCCCGAACAAATTCAGTCCAGGTCACCGGCACCCAGGAGGT 780
Qy 2771 GGTCTAGCACCAAGCAGGGCCAGCCCTTACACATCCGTGTGACACAGATCATGAGAA 2830
Db 781 AGTCCTTTCACCAAGCAGGGCCAGCCCTTGCATATTCGAGTGACACAGACCAAGAGAA 840
Qy 2831 CAGCACTGCCACCTCGAGATCAAGCCCCCACATCTGAAGAGTTTTTCTCTAGTACCAC 2890
Db 841 CAGCACTGGACTTTGGAGATAACAAGCCCCGACATCTGAAGAAATTTTTTCTAGTACCAC 900
Qy 2891 CGTCATTCCTACTCTTAGGCCAACCCAGAAACCAAGATAAACCATTTATTCATCACCCCAATGT 2950
Db 901 TGTCAATTCCTACTCTTAGGGATCAGAAAAACCAAGATAAACCATTTATTCATCACCAACGT 960
Qy 2951 CATGTCGAAAAAGCCCAAGTGAGATCCTACTCTCGGCCAGAACGAGCCATGTCCCC 3010
Db 961 TATGCTCAAAAAACAAAAAGTGGAGATACTACTCTTGGCCAGAACGAGCCATGTCCCC 1020
Qy 3011 TGTCAAGATTACTACTATTTCAGAGAGAGAGCCCGGAAGGTGGAAGGAGCGCCTTTGC 3070
Db 1021 AGTCACAATTACTACATTTTCCAGAGAGAAAGACTCCAGAAAAGTGAAGAGGCGCATTTGC 1080


```
QY . 536 GGAAGAAGCAGAGGAGACGTACCGCCGCGCATGCTAGAGCAGCTGCTGCTGGTGAGAAAGTG 595
Db 727 GGAAGAACAGAAAGAAACCTACCGGCGCATGCTAGAGCAGCTGTTGCTGGCCGAGAAAGTG 786
QY 596 TCACAGGCGCACCGTGTACGAGCTGGAGAACGAGAACGACAAAGCACACACTGACTACATGAA 655
Db 787 TCATAGGCGCACCGTATACGAGTTAGAGAACGAGAACGATAAACACACTGACTACATGAA 846
QY 656 CAAGAGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGAGAGAGTTGAAAAAGCTCCT 715
Db 847 CAAGAGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGGGAGAGGTTAAAAAAGCTCCT 906
QY 716 TGAACAAGAAAAAGCTTACCAAGCCCGCAAGAAAAAGGAAAAACGCTAAGCGGCTCAACAA 775
Db 907 TGAACAAGAAAAAGCTTATCAAGCCCGCAAGAAAAAGGAAAAATGCTAAACGACTCAATAA 966
QY 776 ACTTCGAGATGAGCTTGTGAAGCTCAAGTCTCGCCCTCATGTTGGTGGACGAGAGCA 835
Db 967 ACTAAGAGATGAGCTTGTAAACTCAATCCTTTGCACCTCATGCTGGTGGATGAAAGACA 1026
QY 836 GATGCACATCGAGCAACTGGGCCCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAAGCT 895
Db 1027 AATGCACATTGAACAACTTGGCCTGCAAGCCAGAAAGTACAGGATCTTACTCAGAAAGCT 1086
QY 896 GAGGAGGAGGAAGAAAAAACTCAAAGCGGTCACTTACAAATCCAAAGGAGACCCGCAAGAA 955
Db 1087 GAGGGAAGAAAGAGAGAGCTCAAGGCCATTAATCCAAATCCAAAGAGACAGACAGAA 1146
QY 956 GCTGCTCAAGTTAGAGTGAGCTTCAAGACAAAGGCGCTGAGGTTTTTCCAGGAGCAAGA 1015
Db 1147 ATTGCTCAAGTTAGAGTGAGCTTTGAACACAAAGGCTTCGAGGTTTTTCAAGAGCATGA 1206
QY 1016 AGAGATGAACCGCAAAATTTGGCGAATCAAGAAATCTCACAAACGGGCAACTTCGACTCAAACT 1075
Db 1207 AGAGATGAACCGCTAAACTGGCTTAATCAAGAGTCTCAAAATAGGCAACTTAGACTCAAGCT 1266
QY 1076 GGTGGCTTATCGCAAAAGGATTGAGGAGTGGAAGAGACCAATAAAAGCCTTCAGAAAGGC 1135
Db 1267 GGTGGCTTAACCCAAAGAAATCGAGGAGCTAAGAGAGACCAACAAAATCTGCAGAAAGGC 1326
QY 1136 AGAGGAAGAGACTCCAGGAGCTGAGAGAGAAAAATTCGCAAAAGGGGAATGTGAAACTCCAG 1195
Db 1327 AGAGGAAGAACTTCAAGAAATTAAGAGATAAAATTTGCCAAAGGAGAAATGTGGAAACTCTAG 1386
QY 1196 TCTATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTGTGAGATGGAGGGCAAGGATGA 1255
Db 1387 CCTCATGGCAGAAAGTGAAAAATCTTCGAAAGCGTGTGCTTGAAATGGAAGGTAAAGATGA 1446
QY 1256 AGAGATCACGAAGACCGAGGCCAGTGCAGGAGCTGAAGAAAGAGCTCCAAAGAGGAAGA 1315
Db 1447 GGAGATCACTAAACTGAATCCAGTGTAGGGAATTGAGGAAGAAGCTGCAAGAGAGGAAGA 1506
QY 1316 ACACCAAGCAAGGAACCTTAAGACTAGAAAGTGGAAGCTGCAGAAAGAGGATGTCTGAGCT 1375
Db 1507 ACACCATAGTAAGGAGTTCAGACTTGAAGTTGAGAAGCTACAGAAAGAGAAATGTCTGAACT 1566
QY 1376 GGAGAAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGCACCAGCTCCATCTGAA 1435
Db 1567 AGAGAAATTTGGAAGAAGCATTTAGCAAGAGTAATCTGAGTGCACCCAGCTACATTTAA 1626
QY 1436 CCTGGAGAAAGGAGAAAGAACCTTAACCAAGACCTTGCTGAAACGAGCTGGAGGTGGTCAAGAG 1495
Db 1627 TCTGGAGAAAGAAAGAACTTAACCAAGACCTTGCTAAATGAAATTGAGGTTGGTCAAGAG 1686
QY 1496 TCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAAGCGCGAGTTAAGCCTCAA 1555
Db 1687 TCGAGTTAAAGAAATTGGAATGTTCTGAAAGTAGATTGGAAGAGGCTGAAATTAAGCCTAAA 1746
QY 1556 AGATGACCTTACAAAGCTGAAGTCTTCACTGTGATGCTGGTGGATGAGAGG 1607
Db 1747 AGATGATCTTACCAAGTTGAAGTCAATTTACCGTGTGCTGGTTGATGAAAGG 1798
```

```
RESULT 5
US-10-309-851-13
; Sequence 13, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
; CURRENT APPLICATION NUMBER: US/10/309,851
; CURRENT FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 13
; LENGTH: 3415
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: GIP130b
; NAME/KEY: CDS
; LOCATION: (12)..(3410)
; OTHER INFORMATION:
US-10-309-851-13
```

```
Query Match 18.1%; Score 790.2; DB 15; Length 3415;
Best Local Similarity 55.2%; Pred. No. 1.7e-202;
Matches 1680; Conservative 0; Mismatches 1328; Indels 33; Gaps 6;

QY 286 CTGTGGAGTTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGGAGTTGC 345
Db 202 CAGAAGACCTCTCAAGAGATGACCTGTTATTTCTCCTCAGCATTTCTGGAGGGAGAACTGC 261
QY 346 AGGCTCGAGAAGATGTCATCCACATGCTGAGGACAGAGAAAAACCAAGCCCCGAGTTCTGG 405
Db 262 AGGCTCGAGATGAGTCAAGGCAATTTAAAGGCTGAAAAAATGGACCTGGCTTTGCTGG 321
QY 406 AGGCACACTATGGATCTGCAGAAACCTGAGAAAAAGTGTTCGGGTCTCTGCACCGAGATGCCA 465
Db 322 AAGCTCAGTATGGGTTGTCACTCCAAAAAAGGTGTAGAGGCTCTCCAGAGAGATGCTT 381
QY 466 TCCTTGCTCAAGAGAAGTCCATAGGAGAAGACGTACCCGCGCATGCTAGAGCAGCTGCTGCTGG 525
Db 382 TTCAAGCGAAATCTACCCCTTGGCAGGAGGACATCTATGAGAAACCAATGAATGAGTTGG 441
QY 526 ACAGACTGGAGGAAAAAGCAGAAAGGAGACGTACCCGCGCATGCTAGAGCAGCTGCTGCTGG 585
Db 442 ACAAGTTGTGAAAAAACATAAAGAAATCTTACAGACGAATCCTGGGACAGCTTTTACTGG 501
QY 586 CTGAGAAAGTGTCAAGGCGCACCCGTGTACGAGCTGGAGAAACGAGAAAGCACAACGCACTG 645
Db 502 CAGAAAAATCCCATAGGCAACCATATTTGGAGTTGGAGGAGAAAGAAAAAGAGAAAAAG 561
QY 646 ACTACATGAACAAGAGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGAGAGAGTTGA 705
Db 562 AATAACATGGAGAAGAGTGATGAATTCATATGCCCTACTAGAACACAGGAATGTGAAAAGATTAA 621
QY 706 AAAAGCTCCTTGAAACAAGAAAAAGCTTACCAAGCCCGCAAGAAAAAGGAAAAACGCTAAGC 765
Db 622 AGAAGCTAATTGATCAAGAAATCAAGTCTCAGGAGGAGAAAGGAGCAAGAAAAAGGAGAAAA 681
QY 766 GGCTCAACAACTTCGAGATGAGCTTGTGAAGTCAAGTCCCTCGCCCTCATGTTGGTGG 825
Db 682 GGTTCACCACCCCTGAAAGAGGAGCTGACCAAGCTGAAGTCTTTTGTGTTGATGGTGGTGG 741
QY 826 ACGAGAGGCGAGATGCACATCGAGCAACTGGGCTGCAGAGTCAAGAAAGTCCAGGACCTCA 885
Db 742 ATGAACAGCAAAAGGCTGACGGCACAGCTCACCTTCAAGAGACAGAAAAATCCAAGAGCTGA 801
QY 886 CTCAGAAGCTGAGGGAGGAGGAGAAAAAACTCAAAAGCGGTCACTTACAAATCCAAGGAAG 945
```


Db 1639 AAAGGGCGCTCAAGTCCAAAACCGATGTAGAAGAAAAGATGTACAGCGTAACCAAGGAGA 1698
QY 1786 GGGATGAGCTGATGGGTAAACTGAGGAGCGAAGAAAGGTCTCTGTGAACAGCTGCA 1845
Db 1699 GAGATGATTTAAAAACAAATTGAAAGCGGAAGAGAAAGAAATGATCTCTGTCAA 1758
QY 1846 GTGTAGACTTACTAAAGACGGCTTGATGGCATAGAGGAGGTAGAAAGGGAATAAACC 1905
Db 1759 GAGTTAATATGTTGAAAAATAGGCTTCAATCATTTGGAAGCAATTGAGAAAGATTCTCTAA 1818
QY 1906 GAGGTAGGTCGTGCAAGGGTCTGAGTTCAAC-----TGCCCCGGAAGACAATA 1953
Db 1819 AAAACAAATTAAATCAAGACTCTGGGAAATCCACAACAGCATTTACCAAGAAAAACAATA 1878
QY 1954 AGATCAGAGAACTAACGCTTGAAATCGAGAGACTGAAGAAAAAGGCTCCAGCAGTTGGAGG 2013
Db 1879 AGATTAAAGGAGCTCTCTCAAGAAGTGGAAAGACTGAAAGCTGAAAGCTAAAGGACATGAAAG 1938
QY 2014 TGGTGGAGGGGACTTGATGAAGACCGGAGGACGAAATATGACCAAGTTGGAGCAGAAGTTCA 2073
Db 1939 CCATTGAGGATGACCTCATGAAAAACAGAAAGATGAATATGAGACTCTAGAACGAAAGTATG 1998
QY 2074 GAACCGAGCAGGATAAGSCAAACTTCCTCTCCCGAGAGCTCGAGGAAATCAACACCAAA 2133
Db 1999 CTAATGAACGAGACAAAGCTCAATTTTATCTAAAGAGCTAGAACATGTTTAAATGGAAC 2058
QY 2134 TGGCCAAGCACAAAGCCATAGAGAAAGGGGAGGCGGTGAGCCAGGAAGCCGAACTGCGAC 2193
Db 2059 TTGCTAAGTACAAAGTTAGCAGAAAAAGACAGAGACC--AGCCATGAACAATGGCTTTTCA 2115
QY 2194 ACAGGTTTCGGCTGGAGGAGGCTAAAAGTCGTGATTTACAGGCCGAGGTGACGCTCTCA 2253
Db 2116 AAAGGCTTCAAGAAAGAAAGCTAAGTCAGGGCACCTCTCAAGAGAAAGTGGATGCATTAA 2175
QY 2254 AGGAGAAGATCCACGAGCTGATGAACNAGGAAGCAGCTGTCTCAGCTCCAAGTCGACT 2313
Db 2176 AAGAGAAAAATTCATGATACATGGCAACTGAAGACCTAAATATGTACCTCCAGGAGATC 2235
QY 2314 ATTCGGTCTCTCAGCAAAAGATTATGGAAGAAAGAACTAAGAACAAAGACATGGGGAGG 2373
Db 2236 ACTCAGTCTCTGCAAAAAAACTAAATCAACAAGAAAAACAGGAACAGAGATTTAGGAAGAG 2295
QY 2374 AGGTCTCAATCTGACCAAGGAGCTAGAGCTTTTCAAGCGCTACAGCCGAGCTCTCAGGC 2433
Db 2296 AGATTGAAAACTCACTAAGGAGTTAGAGAGGTACCGGCATTTTCAGTAAGACCTCAGGC 2355
QY 2434 CGAGTGGGAACGGCCGAAGGATGGTGGACGTCCTGTGGCTCCACTGGGGTGCAGACCG 2493
Db 2356 CTAGTCTCAATGGAAGAAGAAATTTCCGATCCTCAAGTATTTTCTAAAGAAAGTTCAAGACAG 2415
QY 2494 AGCGGTGTGCGGGATGCTGCGGAGGAGGAGACCCCGCTGTGTTTCATTTCGCAAAATCCT 2553
Db 2416 AAGCAGTAGACAATGAACCACTGATTACAAGAGCCTCATCTCTGGAACGTGCAGTCA 2475
QY 2554 TCCAGGAGGAAAAATCACATCATGAGTAATCTTCGACAGGTAGGCCCTGAAGAAACCCATGG 2613
Db 2476 TCAATGGTCAGTTATATGAGGAGGTGAGAATCAAGACGAGGACCCCTAATGAT-----G 2529
QY 2614 AACGGTCCCTCGCTCGACAGGATATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGT 2673
Db 2530 AGGGATCTGTGCTCTTCAATGCAGCCAGTCTACTCCAATGTCTGTTAAACAGAAAGC 2589
QY 2674 CTTGGATTCTTGGATGAGAAAAAGAGAAAAACGGTCTTCCACTCCGAGGAGAAAAAGGC 2733
Db 2590 TATGGATTCCCTGGATGAAATCCAAGGAGGGCCATCTTCAGAATGGAATAATGCAA--A 2646
QY 2734 CCAGGCCAAACCAAGGTGCAGGGCACCCCGGAGCTGGTCTCTAGCACCAACCAAGCAGGGCC 2793
Db 2647 CTAAACCCCAATGCCAACTTTGTGCAACCTGGAGATCTAGTCTTAAGCCACACACCTGGGC 2706
QY 2794 AGCCCCCTACACATCCCGTGTGACACCAGATCATGAGAACAGCACTGCCACCCCTGGAGATCA 2853

Db 2707 AGCCACTTTCATATAAAGGTTACTCCAGAGCACCATGTACAAAAACACAGCCACTCTTGAATCA 2766
QY 2854 CAAGCCCCACATCTGAAG-----AGTTTTTCTTAGTACCAACCGTCAATTCCTACCTTAG 2907
Db 2767 CAAGTCCAACCCACAGAGAGTCTCACTCTTACACGAGTACTGCAGTGATACCGAACTGTG 2826
QY 2908 GCAACCAGAAACCAAGAATAAACCATTTATTCATCACCCCAATGTCTATGTCGCAAAAAGCCCA 2967
Db 2827 GCACGCCAAAGCAAGGATAACCATCTCTCCAAAACGCCCTCCATAACACCAGTAAAGTCCA 2886
QY 2968 AAA--GTGCAGATCCTACTCTCGGCCCCAGAACGAGCCCATGTCCCCCTGTACAGATTACTA 3024
Db 2887 AAACCTCTACCGAAGACCTCATGAATTTAGAACAAAGGCATGTCCCCCAATTACCATGGCAA 2946
QY 3025 CTATTTCCAGAGAGAAGAGCCCCGGAAGGTGGRAAGAGCGCCTTTGCCGACAGGCTTCAT 3084
Db 2947 CTTTGGCCAGAGCACAGACCCCAAGAGTCTTTGTGGTTCTTAACTCCAGAAAGGACAATGT 3006
QY 3085 CCCCCATCCAAATCATGACGGGTGTCAACATCTGCAGACTCCCACCTGAAATCGCTGTCTCTC 3144
Db 3007 CCCCTATTCAAGTTTGGCTGTGACTGGTTCAGTACTCTCTTGAGCAGGACGCTCCC 3066
QY 3145 CTGAATCTCAGGAAGTGCCTATGGGAAGGACTATCTCTCAAAGTCAACCCCGGAAAAACAAA 3204
Db 3067 CAGAACCAACAGAAATCAGTGCCNAGCATGCGATATTTCAGAGTCTCCCCCAGACCGGCAGT 3126
QY 3205 CTGTTCCAGCCCCCGTGGGAAGTACAACCTCCAATGCTAAATATCATCACACGGAAGACA 3264
Db 3127 CATCATGGCAGTTTCAGCGTTCAAACACAGCAATAGCTCAAGTGTGATACTACTGAGGATA 3186
QY 3265 ATAAAAATTCACATTCACCTGGGTTCTCAGTTTAAAGCGATCT 3305
Db 3187 ATAAAAATCCACATTCACCTTAGGAAGTCCCTTACATGCAAGCT 3227

RESULT 7

US-10-309-851-11
; Sequence 11, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
; CURRENT APPLICATION NUMBER: US/10/309,851
; CURRENT FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 11
; LENGTH: 3430
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: GIP130a
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9)..(3413)
; OTHER INFORMATION:
US-10-309-851-11

Query Match 18.1%; Score 788.6; DB 15; Length 3430;
Best Local Similarity 55.2%; Pred. No. 4.7e-202;
Matches 1679; Conservative 0; Mismatches 1329; Indels 33; Gaps 6;
QY 286 CTGTGGAGTTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGAGTTGC 345
Db 199 CAGAAGACCTCTCAAGAGATGACCTGTTATTCTCCTCAGCATTTCTGAGGAGAACTGC 258
QY 346 AGGCTCGAGAAGATGTCCACATGTCTGAGGACAGAGAAAAACCAAGCCGAGGTTCTGG 405
Db 259 AGGCTCGAGATGAGGTATAGGCATTTTAAAGGCTGAAAAAATGGACCTTGGCTTGTGG 318

Qy	406	AGGCACACTATGGATCTGCAGAACTTGAGAAAGTGTCTTCGGGTCTCTGCACCGAGATGCCA	465
Db	319	AAGCTCAGTATGGGTTTGTCACCTCCAAAAGGTGTTAGAGGCTCTCCAGAGAGATGCTT	378
Qy	466	TCCTTGCTCAAGAGAAGTCCATAGGAGAGACGCTCTATGAGAAACCTATCTCAGAGCTGG	525
Db	379	TTCAAGCGAAATCTACCCCTTGGCAGGAGGACATCTATGAGAAACCAATGAATGATTGG	438
Qy	526	ACAGACTGGAGGAAAAAGCAGAAAGGAGACGTACCGCGCATGTCTAGAGCAGCTGCTGCTGG	585
Db	439	ACAAAGTTGTGGAAAAACATAAAGAAATCTTACAGACGAAATCCTGGGACAGCTTTTAGTGG	498
Qy	586	CTGAGAAGTGTACAGGCGCACCGTGTACGAGCTGGAGAACGAGAAAGCAACAGCACACTG	645
Db	499	CAGAAAAATCCCGTAGGCAAAACCATATTGGAGTTGGAGGAAGAAAGAGAAAAACATAAG	558
Qy	646	ACTACATGAACAAGACCGACGACTTCACCAACCTGCTGGAGCAGGAGCGAGAGAGGTTGA	705
Db	559	AATACATGGAGAAAGATGATGAATTCATATGCCTACTAGAACAGGAATGTGAAAGATTAA	618
Qy	706	AAAAGCTCCTTGAAACAAGAAAAAGCTTTACCAAGCCCGCAAAAGAAAAGGAAAAACGCTAAGC	765
Db	619	AGAAGCTTAATTGATCAAGAAATCAAAGTCTCAGGAGGAGAGGACCAAGAAAAAGGAGAAA	678
Qy	766	GGCTCAACAAAACCTTCGAGATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCATGTTGGTGG	825
Db	679	GGGTCAACCACCTTGAAAGAGGAGCTGACCAAGCTGAAGTCTTTTGTCTTGATGGTGGTGG	738
Qy	826	ACGAGAGGCAGATGCACATCGAGCAACTGGGCCCTGCAGAGTCAGAAAGTCCAGGACCTCA	885
Db	739	ATGAACAGCAAAAGGCTGACGGCACAGCTCACCCCTTCAAGACAGAAAAATCCAAGAGCTGA	798
Qy	886	CTCAGAAGCTGAGGAGGAGGAAGAAAACCTCAAAGCGGTCACTTACAAATCCAAAGGAAG	945
Db	799	CCACAAATGCAAGGAAACACATACCAAACTAGCCCTTGCTGAAGCCAGAGTTCAGGAGG	858
Qy	946	ACCGCCAGAAGCTGCTCRAAGTTAGAAGTGGACTTCGAACACAAAGGCCTCGAGGTTTCCC	1005
Db	859	AAGAGCAGAAGGCCAACCCAGACTAGAGAAAGAACTGCAAAACGCAGACCACAAAGTTTCACC	918
Qy	1006	AGGAGCAGCAAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCACAAACGGCAACTTC	1065
Db	919	AAGACCAAGACACAAATTATGGCGAAGCTCACCAATGAGGACAGTCAAAATCGCCAGCTTC	978
Qy	1066	GACTCAAACTGGTTGGCTTATCGCAAGGATTGAGGAGCTGGAAGACCAATATAAGGCC	1125
Db	979	AACAAAAGCTGGCAGCACTCAGCCGGCAGATTGATGAGTTAGAAGAGACAAACAGGTCTT	1038
Qy	1126	TTCAAGAAAGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAATTGCCAAAGGGGAATGTG	1185
Db	1039	TACGAAAAAGCAGAGAGAGCTGCAAGATATAAAGAAAAAAATCAGTAAGGGAGAAATATG	1098
Qy	1186	GAAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTGTCTTGAGATGGAGG	1245
Db	1099	GAAACGCTGGTATCATGGCTGAAGTGGAAAGAGCTCAGGAAACGTGTGTAGATATGGAAG	1158
Qy	1246	GCAAGGATGAAGAGATCACGAAGACCGAGGCCCCAGTSCCGGGAGCTGAAGAAAGAGCTCC	1305
Db	1159	GGAAAGATGAAGAGCTCATAAAAATGGAGGAGCAGTGCAGAGATCTCAATAAGAGGCTTG	1218
Qy	1306	AAGAGGAAGAAACACCCACAGCAAGGAACTTAGACTAGAAAGTGGAGAAAGCTGCAGAAAGGA	1365
Db	1219	AAAGGAGACGTTACAGAGTAAAGACTTTAAACTAGAGGTTGAAAAACTCAGTAAAGAA	1278
Qy	1366	TGCTCAGCTGGAGAGCTGGAGGAAGGTTTCAGCCGAGTAAAGTGGAAATGCACCCAGC	1425
Db	1279	TTATGGCTCTGGAAAAAGTTAGAAGACGCTTTCAACAAAAGCAACRAAGATGCTACTCTC	1338
Qy	1426	TCCATCTGAACCTGGAGAAGGAGAAACCTTAACCAAGACCTGCTGAACGAGCTGGAGG	1485
Db	1339	TGAATGCAATTTAGAAAAAAGAAAGGATGACCAACAAGCAGTTGTCTCAAGAACTGGAGA	1398

Qy	1486	TGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAAGGCCGAGT	1545
Db	1399	GTTTAAAAAGTAAGGATCAAGAGCTAGAAGCCATTGAAAGTCGGCTAGAAAAGACAGAAAT	1458
Qy	1546	TAAGCCTCAAAAGATGACCTTACAAAAGCTGAAAGTCCCTTCACTGTGATGCTGTTGGATGAGA	1605
Db	1459	TCACCTAAAAAGAGGATTTAACTTAACTGAAAACATTAACTGTGATGTTGTAGATGAAC	1518
Qy	1606	GAAAAAATATGATGGAGAAAAATAAAGCAAGAAAGAGAGAAAAGTGGATGGGTTGAAATAAA	1665
Db	1519	GGAAAACAATGAGTGAAAAATTTAAAGAAAACTGAAGATAAATTACAAGCTGCTTCTTCTC	1578
Qy	1666	ACTTTAAGGTGGAGCAGGGAAAAGTCATGGATGTGACGGAAAAAGCTAATCGAGGAAAGCA	1725
Db	1579	AGCTTCAAGTGGAGCAAAATAAAGTAAACAAACAGTTACTGAGAAAGTTAATTGAGGAAACTA	1638
Qy	1726	AGAAGCTTTTAAAACTCAAATCTGAAATGGAGGAAAAGGAGTACAGTCTGACAAAAGGAGA	1785
Db	1639	AAAGGCGCTCAAGTCCAAAAACCGATGTAGAAGAAAAAGATGTACAGCGTAACCAAGGAGA	1698
Qy	1786	GGGATGAGCTGATGGGTAAACTGAGGAGCGAAGAAAGAAAGGTCTCTGTGAACTGAGCTGCA	1845
Db	1699	GAGATGATTTAAAAACAATTTGAAAGCGGAAAGAGAGAAAAGGAATGATCTCCTGTCAA	1758
Qy	1846	GTGTAGACTTACTTAAAGAAAGCGGCTTGATGGCATAGAGGAGGTAGAAAAGGAAAATAAACCC	1905
Db	1759	GAGTTAAATATGTTGAAAAATAGGCTTCAATCATTTGGAAGCAATTGAGAAAAAGATTTCCTAA	1818
Qy	1906	GAGTPAGGTCTGCAAGGGGTCTGAGTTTCACTGAGTTCACCTGAGGCGGAAAGACAATA	1953
Db	1819	AAAACAAAATTAATCAAGACTCTGGGAAATCCACAACAGCATTTACACCAAGAAAAACAATA	1878
Qy	1954	AGATCAGAGAACTAAACGCTTGAATCGAGAGACTGAAGAAAAACGGCTCCAGCAGTTGGAGG	2013
Db	1879	AGATTAAAGGAGCTCTCTCAAGAAAGTGGAAAGACTGAAAACCTGAAGCTTAAAGGACATGAAAG	1938
Qy	2014	TGTTGAGGGGAGCTTGATGAAGACCGGAGGACGAAATATGACCAGTTGGAGCAGAAAGTTCA	2073
Db	1939	CCATTGAGGATGACCTCATGAAAAACAGAAGATGAATATGAGACTCTAGAAACGGAAGGTATG	1998
Qy	2074	GAACCGAGCAGGATAAGGCCAAACTTCTCTCCAGCAGCTCGAGGAAATCAAACACCAAA	2133
Db	1999	CTAATGAACGAGACAAAGCTCAATTTTTTATCTAAAGAGCTAGAACATGTTAAATGGAAC	2058
Qy	2134	TGGCCAAAGCACAAAGCCATAGAGAAAGGGGAGGCGGTGAGCCAGGAAAGCCGAACTGCGAC	2193
Db	2059	TTGCTAAGTACAAGTTAGCAGAAAAAGACAGAGACC--AGCCATGAACAAATGGCTTTTCA	2115
Qy	2194	ACAGGTTTCGGCTGGAGGAGGCTTAAAGTCTGTGATTTACAGGCGCAGGTCAGGCTCTCA	2253
Db	2116	AAAGGCTTCAAGAAAGAAAGCTAAAGTCAGGGCACCTCTCAAGAGAAAGTGGATGCATTA	2175
Qy	2254	AGGAGAAAGATCCAGCAGCTGATGAACAAAGGAAGCAGCTGTCTCAGCTCCAAGTCGACT	2313
Db	2176	AAGAGAAAATTCATGAATACATGGCAACTGAAGACCTAATATGTACACCTCCAGGGAGATC	2235
Qy	2314	ATTCCGTCCTTCAGCAAAAGATTTATGGAAGAGAAAACCTAAGAACAAAGAACATGGGGAGGG	2373
Db	2236	ACTCAGTCTCTGCAAAAAAAACTTAAATCAACAAGAAAAACAGGAACAGAGATTTAGGAAGAG	2295
Qy	2374	AGGTCCTCAATCTGAACCAAGGAGTAGAGCTTTTCCAAAGCGCTACAGCCGAGCTCTCAGGC	2433
Db	2296	AGATTGAAAAACCTCACTAAGGAGTTAGAGAGGTACCGGCATTTTCAGTAAGAGCCTCAGGC	2355
Qy	2434	CGAGTGGAAACCGCCGAAGGATGGTGGACGTGCTGTGGCCTCCACTGGGGTGCAGACCG	2493
Db	2356	CTAGTCTCAATGGAAGAAGAAATTTCCGATCCCTCAAGTATTTTCTTAAAGAAAGTTCAAGACAG	2415
Qy	2494	AGGCGGTGTGGGGGATGCTGCGGAGGAGGAGACCCCGGCTGTGTTCATTTCGCAAAATCCT	2553
Db	2416	AAGCAGTAGACAATGAACCACTGATTACAAGAGCCTCATTCCTCTGGAACGTGCAGTCA	2475
Qy	2554	TCCAGGAGGAAAAATCATCATGAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCATGG	2613

```
Db 2476 TCAATGGTCAGTTATATGAGGAGAGTGAGAATCAAGACGAGGACCCCTAATGAT-----G 2529
QY 2614 AACGGTCCTCGTCTCGACAGGTATCCCCAGCAGCGAATGAGCTCACCATGAGGAAGT 2673
Db 2530 AGGGATCTGTGCTGTCCTTCAAAATGCAGCCAGTCTACTCCATGTCTGTGTTAACAGAAAGC 2589
QY 2674 CTGGGATTCCTTGGATGAGAAAAAGAGAAACCGTCTTCCACTCCGCGAGGAGAAAGGC 2733
Db 2590 TATGGATTCCTGGATGAATCCAAGGAGGCCATCTTCAGAATGGAATAATGCAA---A 2646
QY 2734 CCAGGCCAAACCAAGGTGCAGGCCACCCCGGGAGCTGGTCTTAGCACCAAGCAGGGCC 2793
Db 2647 CTAAACCCCAATGCCAACTTTGTGCAACCTGGAGATCTAGTCTTAAGCCACACACCTGGGC 2706
QY 2794 AGCCCTTACACATCCGTGTGACACACCAGATCATGAGAACAGCACTGCCACCCCTGGAGATCA 2853
Db 2707 AGCCACTTCATATAAAGGTTACTCCAGACCATGTACAAAACACAGCCACTCTTGAAATCA 2766
QY 2854 CAAGCCCCACATCTGAAG-----AGTTTTTCTCTAGTACACCGTCAATTCCTACCTTAG 2907
Db 2767 CAAGTCCAACCACAGAGAGTCTCACTCTTACACGAGTACTGCAGTGATACCGAACTGTG 2826
QY 2908 GCAACCCAGAAACCAAGAAATAACCATATTCCATCACCCAAATGTATGTCGCAAAAGCCCA 2967
Db 2827 GCACGCCAAAGCAAGGATAACCATCTCTCCAAAACGCCCTCCATAACACCAGTAAAGTCCA 2886
QY 2968 AAA---GTGCAGATCCTACTCTCGGCCCCAGAACGAGCCATGTCCCCTGTCAAGATTACTA 3024
Db 2887 AAACCTCTACCGAAGACCTCATGAATTTAGAACAAAGGATGTCCCCCAATTACCATGGCAA 2946
QY 3025 CTATTTCCAGAGAGAAGAGCCCGGAAGGTGGAAGAGCGCCTTTGCCGACAGGCCCTGCAT 3084
Db 2947 CCTTTGCCAGAGCACAGACCCAGAGTCTTGTGTTCTCTAACTCCAGAAAGGACAATGT 3006
QY 3085 CCCCCATCCAAATCATGACGTGTCAACATCTGCAGTCCCACTGAAATCGTGTCTCTC 3144
Db 3007 CCCCTATTACAGTTTTTGGCTGTGACTGGTTTCAGCTAGTCTCTCCTGAGCAGGACGCTCCC 3066
QY 3145 CTGAATCTCAGGAAGTGCCTATGGGAAGGACTATCCTCAAAAGTCACCCCGGAAACAAA 3204
Db 3067 CAGAACCAACAGAAATCAGTGCCCAAGCATCGGATATTTCAGAGTCTCCCCAGACCGGCAGT 3126
QY 3205 CTGTTCCAGCCCCCGTGGGAAGTACAACTCCAATGCTAATATATCATCACCCAGGAAGACA 3264
Db 3127 CATCATGGCAGTTTCAGCGGTTCAACACAGCAATAGCTCAAGTGTGATAACTACTGAGGATA 3186
QY 3265 ATAAAAATTCACATTACCTGGGTTCTCAGTTTAAAGCGATCT 3305
Db 3187 ATAAAAATCCACATTCACTTAGGAAGTCTCTTACATGCAAGCT 3227
```

```
RESULT 8
US-10-309-851-9
; Sequence 9, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
; CURRENT APPLICATION NUMBER: US/10/309,851
; CURRENT FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 3998
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (473)..(2767)
```

```
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: GIP90
US-10-309-851-9

Query Match 17.8%; Score 777.6; DB 15; Length 3998;
Best Local Similarity 55.2%; Pred. No. 5e-199;
Matches 1679; Conservative 0; Mismatches 1329; Indels 34; Gaps 7;

QY 286 CTGTGGAGTTATCCAAGGAGGACCTCATCCAGTCTCCTGAGTATCATGGAAGGGAGTTGC 345
Db 663 CAGAAGACCTCTCAAGAGATGACCTGTTATTTCTCCTCAGCATTTCTGGAGGGAGACTGC 722
QY 346 AGGCTCGAGAAGATGTCTATCCACATGCTGAGGACAGAGAAAAACCAAGCCCGAGGTTCTGG 405
Db 723 AGGCTCGAGATGAGGTTCATAGGCATTTTAAAGGCTGAAAAAATGGACCTGGCTTTGCTGG 782
QY 406 AGGCACACTATGGATCTGCAGAACCTGAGAAAAGTCTTCGGGTCTCTGCACCCGAGATGCCA 465
Db 783 AAGCTCAGTATGGGTTTGTCACTCCAAAAAAGGTGTAGAGGCTCTCCAGAGAGATGCTT 842
QY 466 TCCTTGTCTCAAGAGAAATCCATAGGAGAACGTCGTCTATGAGAAAACCTATCTCAGAGCTGG 525
Db 843 TTCAAGCGAAATCTACCCCTTGGCAGGAGGACATCTATGAGAAAAACCAATGAATGAGTTGG 902
QY 526 ACAGACTGGAGGAAAAAGCAGAAAGGAGACGTACCCGCGCATGCTAGAGCAGTCTGCTGCTGG 585
Db 903 ACAAAGTTGTGAAAAACATAAAGAATCTTACAGACGAATCTCTGGGACAGCTTTTACTGG 962
QY 586 CTGAGAAAGTGTACAGCGCCACCGTGTACGAGCTGAGAAACGAGAACGACACACTG 645
Db 963 CAGAAAAATCCCGTAGGCAAAACCATATTGGAGTTGGAGGAAGAAAGAGAAAAACATAAAG 1022
QY 646 ACTACATGAACAAGAGCGACGACTTCAACCAACCTGCTGGAGCAGGAGCGAGAGAGGTTGA 705
Db 1023 AATACATGGAGAAGAGTGATGAATTCATATGCTTACTAGAACAGGAATGTGAAGATTA 1082
QY 706 AAAAGCTCCTTGAAACAGAAAAAGCTTACCAAGCCCGCAAGAAAAAGGAAACCGTAAAGC 765
Db 1083 AGAAGCTAATTGATCAAGAAATCAAGTCTCAGGAGAGAGGAGCAAGAAAAAGGAGAAAA 1142
QY 766 GGCTCAACAACTTCGAGATGAGCTTGTGAAGTCAAGTCTTTCGCCCTCATGTTGGTGG 825
Db 1143 GGGTCACCACCCCTGAAAGAGGAGCTGACCAAGCTGAAGTCTTTTGTCTTGTATGTTGGTGG 1202
QY 826 ACGAGAGGCAGATGCACATCGAGCAACTGGGCCCTGCAGAGTCCAGAAAGTCCAGACCTCA 885
Db 1203 ATGAACAGCAAAAGGCTGACGGCACAGCTCACCCCTTCAAAGACAGAAAAATCCAGAGCTGA 1262
QY 886 CTCAGAAGCTGAGGAGGAGGAAGAAAAACTCAAAGCGGTCACTTACAAATCCAAAGGAAG 945
Db 1263 CCACAAATGCAAGGAAACACATACCAAACTAGCCCTTGTGAAGCCAGAGTTTCAGGAGG 1322
QY 946 ACCGCCAGAAAGCTGCTCAAGTTAGAAAGTGGACTTCGAAACACAAGGCCCTCGAGGTTTCCC 1005
Db 1323 AAGAGCAGAAGGCAACAGACTAGAGAAGGAACTGCAAAACGCAGACCAACAAAGTTTCACC 1382
QY 1006 AGGAGCACGAAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCACAACCCGCAACTTC 1065
Db 1383 AAGACCAAGACACATATTATGGCGAAGCTCACCAATGAGGACAGTCAAAATCGCCAGCTTC 1442
QY 1066 GACTCAAACTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCC 1125
Db 1443 AACAAAAGCTGGCAGCACTCAGCCGGCAGATTGATGATTAGAGAGACAAACAGGTCCTT 1502
QY 1126 TTCAGAAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGGAATGTG 1185
Db 1503 TACGAAAAGCAGAGAGGAGCTGCAAGATATAAAAGAAAAAATCAGTAAGGGAGAAATATG 1562
QY 1186 GAAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCTGCTTGTAGATGGAGG 1245
Db 1563 GAAACGCTGTTATCATGGCTGAAGTGGAAAGAGCTCAGGAAACGTTGCTAGATATATGGAAG 1622
```


; GENERAL INFORMATION:									
; APPLICANT: HELIX RESEARCH INSTITUTE									
; TITLE OF INVENTION: No. US20040005560Ae1 full length cDNA									
; FILE REFERENCE: H1-A0106									
; CURRENT APPLICATION NUMBER: US/10/108,260A									
; CURRENT FILING DATE: 2002-03-27									
; NUMBER OF SEQ ID NOS: 5458									
; SOFTWARE: PatentIn Ver. 2.1									
; SEQ ID NO 1126									
; LENGTH: 3069									
; TYPE: DNA									
; ORGANISM: Homo sapiens									
US-10-108-260A-1126									
Query Match 12.9%; Score 565; DB 16; Length 3069;									
Best Local Similarity 54.2%; Pred. No. 1.8e-141;									
Matches 1297; Conservative 0; Mismatches 1065; Indels 33; Gaps 6;									
QY	932	CAAAATCCAAGGAAGACCCGACGAAGCTGCTCAAGTTAGAAAGTGGACTTCGAAACACAAGGC	991						
Db	159	CAAGTTTCAGGAGGAAGACGAGAAGGCCAACCGACTAGAGAAGGAACCTGCAAAACCGCAGAC	218						
QY	992	CTCGAGGTTTTCCAGGAGCAGCAAGAGATGAACGCCAAATTTGGCGAATCAAGAACTCTCA	1051						
Db	219	CACAAAGTTTCACCAAGACCAAGACACAAATTTATGGCGAAGCTCACCAATGAGGACAGTCA	278						
QY	1052	CAACCGGCAACTTCGACTCAAACTGGTTGGCTTATCGCAAAAGGATTGAGGAGCTGGAAGA	1111						
Db	279	AAATCGCCAGCTTCAACNAAAGCTGGCAGCACTCAGCCGGCAGATTGATGAGTTAGAAGA	338						
QY	1112	GACCAATAAAAGCCTTCAGAAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAATTGC	1171						
Db	339	GACAAACAGGCTCTTTACGAAAAGCAGAAGAGGAGCTGCAAGATATAAAAGAAAATCAG	398						
QY	1172	CAAAGGGGAATGTGGAAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGGT	1231						
Db	399	TAAGGGAGAATATGGAACCGCTGGTATCATGGCTGAAGTGAAGAGCTCAGGAACCGTGT	458						
QY	1232	GCTTGAGATGGAGGGCAAGGATGAAGAGATCACGAAGACCGAGGCCAGTGCCGGGAGCT	1291						
Db	459	GCTAGATATGGAAGGGAAAGATGAAGAGCTCATAAAAATGGAGGAGCAGTGCAGAGATCT	518						
QY	1292	GAAGAAGAAAGCTCCAAGAGGAAGAACAACCCACAGCAAGAAACTTAGACTAGAAAGTGGAGAA	1351						
Db	519	CAATAAGAGGCTTGAAAGCGAGACGTTACAGAGTAAAGACTTTAAACTAGAGGTTGAAAA	578						
QY	1352	GCTGCAGAAGAGGATGTCTGAGCTGGAGAAGCTGGAGAAGCGTTTCAGCCGGAGTAAAGTC	1411						
Db	579	ACTCAGTAAAGAATTATGGCTCTGGAAAAGTTAGAAGACGCTTTCAACAAAAGCAACA	638						
QY	1412	GGAATGCACCCAGCTCCAATCTGAACCTGGAGAAGGAGAAGAACCTTAACCAAGACCTTGCT	1471						
Db	639	AGAATGCTACTCTCTGAATAATGCAATTTAGAAAAAGAAAGGATGACCAACAAGCAGTTGTC	698						
QY	1472	GAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACT	1531						
Db	699	TCAAGAACTGGAGAGTTTAAAAAGTAAAGGATCAAAAGAGCTAGAAGCCATTGAAAAGTCGGCT	758						
QY	1532	GGAGAAGGCCGAGTTAAGCTCAAGATGACCTTTACAAAGCTGAAGTCTTCACTGTGTAT	1591						
Db	759	AGAAAAGACAGAAATTCACCTTAAAAGAGGATTTAACTTAAACTGAAAACATTTAACTGTAT	818						
QY	1592	GCTGGTGGATGAGAGGAAAAATATGATGGAGAAAAATAAGCAAGAAGAGAGGAAAGTGA	1651						
Db	819	GTTTGTAGATGAACGGGAAAAACAATGAGTGAAAAATTTAAAGAAAACTGAAGATAAATTACA	878						
QY	1652	TGGGTTGAATAAAAACTTTAAGGTGGAGCAGGGAAGTCAATGATGTGACGGGAAAAGCT	1711						
Db	879	AGCTGCTTCTTCAGCTTCAAGTGGAGCAAAAATAAAGTAAACAACAGTTTACTGAGAAGTT	938						
QY	1712	AATCGAGGAAAAGCAAGAGCTTTTAAAACTCAAACTCGAAATGGAGGAAAAAGGAGTACAG	1771						
Db	939	AATTGAGGAAACTTAAAGGGCGCTCAAGTCCAAAACCGGATGTAGAAGAAAAAGATGTACAG	998						

QY	1772	TCTGACAAAGGAGAGGATGAGCTGATGGGTAAACTGAGGAGCGGAAGAAGAAAGTCCTG	1831						
Db	999	CGTAACCAAGGAGAGAGATGATTTAAAAAACAATAATTGAAAGCGGAAGAAGAGAAAGGAAA	1058						
QY	1832	TGAACTGAGCTGCAGTGTAGACTTACTATAAGAAAGCGGCTTGATGGCATAGAGAGGTAGA	1891						
Db	1059	TGATCTCCTGTCAAGAGTTAATATGTTGAAAAATAGGCTTCAATCATTTGGAAGCAATTGA	1118						
QY	1892	AAGGGAATAAACCGAGGTAGGTGTCGTCGAAGGGGTCTGAGTTCACC-----TG	1939						
Db	1119	GAAAGATTTCTTAAAAAACAATAATAATCAAGACTCTGGGAAATCCACAACAGCATTACA	1178						
QY	1940	CCCGGAAGACAATAAGATCAGAGAACTAAACGCTTGAATCGAGAGACTGAAGAAACGGCT	1999						
Db	1179	CCAAGAAAAACAATAAGATTAAAGAGCTCTCTCAAGAAGTGGAAAGACTGAAACTGAAGCT	1238						
QY	2000	CCAGCAGTTGGAGTGGTGGAGGGGACTTGATGAAGACCGAGGACCGAATATGACCAGTT	2059						
Db	1239	AAAGGACATGAAAGCCATTGAGGATGACCTCATGAAAAACAGAAGATGAATATGAGACTCT	1298						
QY	2060	GGAGCAGAAGTTCAGAACCGAGCAGGATAAGGCAAACTTCCTCTCCAGCAGCTCGAGGA	2119						
Db	1299	AGAACGAAGGTATGCTAATGAACGAGACAAAGCTCAATTTTATCTAAAGAGCTAGAACA	1358						
QY	2120	AATCAAAACACCAATGGCCAAAGCACAAAGCCATAGAGAAAAGGGAGGCGGTAGCCAGGA	2179						
Db	1359	TGTTAAAAATGGAACCTGCTAAGTACAAAGTTAGCAGAAAAAGACAGAGACC---AGCCATGA	1415						
QY	2180	AGCCGAACCTGCGACACAGGTTTTCGGCTGGAGGAGGCTAAAAAGTCGTGATTTACAGGCCGA	2239						
Db	1416	ACAATGGCTTTTCAAAAAGGCTTCAAGAAGAAAGAGCTAAGTCAGGGCACCTCTCAAGAGA	1475						
QY	2240	GGTGAGGCTCTCAAGGAGAAAGATCCACGAGCTGATGAACAAGGAAGACCAGCTGTCTCA	2299						
Db	1476	AGTGGATGCATTTAAAGAGAAAAATTCATGAATACATGGAAGCTGAAGACCTAATATGTCA	1535						
QY	2300	GCTCCAAGTCGACTATTTCGGTCTTCAGCAAAAGATTATGGAAGAAGAAACTAAGAACAA	2359						
Db	1536	CCTCCAGGGAGATCACTCAGTCTCTGCAAAAAAATAACTAAATCAACAAGAAAAACAGGAACAG	1595						
QY	2360	GAACATGGGGAGGAGGTCTCTCAATCTGACCACAGGAGCTAGAGCTTTCCAAAGCGCTACAG	2419						
Db	1596	AGATTTAGGAAGAGAGATTGAAAAACCTCACTAAGAGTTAGAGAGGTACCGGCAATTTTCAG	1655						
QY	2420	CCGAGCTCTCAGGCCAGTGGGAACGGCCGAAAGGATGGTGGACGTGCCTGTGGCCTCCAC	2479						
Db	1656	TAAGAGCCTCAGGCCTAGTCTCAATGGAAAGAAATTTCCGATCCTCAAGATATTTCTTAA	1715						
QY	2480	TGGGTGCAGACCGAGCGGTGTGCGGGGATGCTGCGGAGGAGAGACCCCGGTGTGTT	2539						
Db	1716	AGAAGTTTCAGACAGAAAGCAGTAGACAATGAACCACTGATTAACAAGAGCCTCATTCCTCT	1775						
QY	2540	CATTGCAAAATCCTTCCAGGAGGAAAAATCACATCATGAGTAAATCTTCGACAGGTAGGCCT	2599						
Db	1776	GGAACGTGCAGTCAATAGGTGAGTTATATGAGGAGAGTGAGAATCAAGACAGGAGACCC	1835						
QY	2600	GAAGAAAACCCATGGAACGGTCTCGGTCTCGACAGGTATCCCCCAGCAGCGAATGAGCT	2659						
Db	1836	TAATGA-----TGAGGGATCCGTGCTGCTTCAAAATGCAGCCAGTCTACTCCATGTCC	1889						
QY	2660	CACCATGAGGAAGTCTTGGATTCTTGGATGAAAAAAGAGAAAAACGGTCTTCTCCACTCC	2719						
Db	1890	TGTTAACAGAAAAGCTATGGATTCCCTGGATGAAATCCAAGGAGGGCCATCTTTCAGAAATGG	1949						
QY	2720	GCAGGAGAAAAGGGCCAGGCCAAACCCAGGTTGACGGGCACCCCGGGAGCTGGTCTTAGC	2779						
Db	1950	AAAAATGCAA---ACTAGACCCAATGCCAACITTTGTGCAACCTGGAGATCTAGTCTCTAAG	2006						
QY	2780	ACCAAAGCAGGGCCAGCCCTTACACATCCCGTGTGACACCAAGATCATGAGAACAGCACTGC	2839						
Db	2007	CCACACACCTGGGGAGCCACTTTCATATAAAGGTTACTCCAGACCATGTACAAAACACAGC	2066						

Db	2483	CGCAGTCATCATGGCAGTTTCAGCGTTCAAACAGCAATAGCTCAAGTGTGATAACTACT	2542
QY	3258	GAAGACAATAAAATTCACATTCACTGGGTTCTCAGTTTAAGCGATCT	3305
Db	2543	GAGGATAATAAAATCCACATTCACTTAGGAAGTCCTTACATGCAAGCT	2590

RESULT 12

```

US-09-873-319-634
; Sequence 634, Application US/09873319A
; Publication No. US20030134324A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; APPLICANT: Kulkarni, Prakash
; APPLICANT: Getzenberg, Robert H.
; APPLICANT: Waga, Iwao
; APPLICANT: Yamamoto, Jun
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic
; TITLE OF INVENTION: Hyperplasia Using Gene Expression Profiles
; FILE REFERENCE: 44921-5029-US
; CURRENT APPLICATION NUMBER: US/09/873,319A
; CURRENT FILING DATE: 2001-06-05
; EARLIER APPLICATION NUMBER: US 60/223,323
; EARLIER FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 755
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 634
; LENGTH: 3025
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134324A1 U534445
US-09-873-319-634

```

Query Match	12.4%;	Score 541.6;	DB 10;	Length 3025;
Best Local Similarity	53.0%;	Pred. No. 3.9e-135;		
Matches 1393;	Conservative	0;	Mismatches 1159;	Indels 76;
				Gaps 8;

Qy	699	AGTTGAAAAGCTCCTTGAACAAGAAAAAGCTTACCAAGCCCGCAAGAAAAAGGAAAAAC	758
Db	18	AGATTAAAGAGCTAATTGATCAAGAAATCAAGTCTCAGGAGGAGAGGACAAGAAAAAG	77
Qy	759	GCTAAGCGGCTCAACAACAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCITTCGCCCTCATG	818
Db	78	GAGAAAAGGGTCACCAACCTGAAAGAGGAGCTGACCAAGCTGAAAGTCTTTGCTTTGATG	137
Qy	819	TTGGTGGACGAGAGGCAGATGCACATCGAGCAACTGGGCCCTGCAGAGTCAGAAAAGTCCAG	878
Db	138	GTGGTGGATGAACAGCAAAGGCTGACGGCAGCTCACCTTCAAAGACAGAAAAATCCAA	197
Qy	879	GACCTCACTCAGAAGCTGAGGGAGGAGGAAGAAAAAACTCAAAGCGGTCACTTACAAATCC	938
Db	198	GAGCTGACCACAAAATGCAAAGGAAACACATACCAAACCTAGCCCTTGTCTGAAGCCAGATT	257
Qy	939	AAGGAAGACCGCCAGAAGCTGCTCAAGTTAGAAGTGGACTTCGAAACACAGGCCCTCGAGG	998
Db	258	CAGGAGGAAGACGAGAAGGCAACCAGACTAGAGAAGGAACTGCAACCGCAGACCACAAAG	317
Qy	999	TTTTCCAGGACGACGAAGAGATGAACGCCAAATTTGGCGAATCAAGAAATCTCAAAACCGG	1058
Db	318	TTTCACCAAGACCAAGACACAATTTATGGCGAAGCTCACCAATGAGGACAGTCAAAAATCGC	377
Qy	1059	CAACTTCGACTCAAACTGGTTGGCTTATCGCAAAAGGATTGAGGAGCTGGAAGAGACCAAT	1118
Db	378	CAGCTTCAACAAAAAGCTGGCAGCACTCAGCGGCAGATTGTAGTTAGAAGAGACAAAC	437
Qy	1119	AAAAGCCTTCAGAAAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGG	1178
Db	438	AGGTCTTTACGAAAAGCAGAAGAGGAGCTCCAAGATATAAAAGAAAAAATCAGTAAGGGA	497
Qy	1179	GAATGTGAAAACCTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTGTCTTGA	1238
Db	498	GAATATGGAACCGCTGGTATCATGGCTGAAGT-----	530

```
QY 2307 GTCGACTATTTCGGTCTCTTCAGCAAAAGATTTCATGGAAGAGAAAACTAAGAAACAAGAACAATG 2366
Db 1593 GGAGATCACTCAGTCTGCAAAAAAATACTAAATCAACAAGAAAAACAGGAACAGAGATTTA 1652
QY 2367 GGGAGGGAGGTCTCTCAATCTGACCAAGAGCTAGAGCTTTCCAAAGCGCTACAGCCGAGCT 2426
Db 1653 GGAAGAGAGATTGAAAACCTCACTAAGGATTAGAGAGGTACCGGCATTTTCAGTAAGAGC 1712
QY 2427 CTCAGGCCAGTGGGAACGSCCGAAGGATGGTGGACGTGCTGTGGCCTCAACTGGGGTG 2486
Db 1713 CTCAGGCCTAGTCTCAATGGAAGAAGAAATTTCCGATCCTCAAGTATTTTCTAAAGAAGTT 1772
QY 2487 CAGACCGAGCGGTGTGCGGGGATGCTGCGGAGGAGGAGACCCCGGTGTGTTTCATTCCG 2546
Db 1773 CAGACAGAAGCAGTAGACAATGAACCACTGATTACAAGAGCCTCATTTCTCTGGAACGT 1832
QY 2547 AAATCCTTCAGGAGGAAAAATCACATCATGAGTAATCTTCGACAGGTAGGCTGAAGAAA 2606
Db 1833 GCAGTCATCAATGGTCAGTTATATGAGGAGAGTGAGAAATCAAGACGAGGACCCCTAATGA- 1891
QY 2607 CCCATGGAACCGTCTCGGTCTCGACAGGTATCCCCCAGCAGCGAAATGAGCTCACCATG 2666
Db 1892 -----TGAGGGATCTGTGCTGTCTTCAATGCGACCCAGTCTACTCCATGCTGTTAAC 1946
QY 2667 AGGAAGTCTTGGATTCTTGGATGAGAAAAAAGAGAAAAACGGTCTCTTCCACTCCGAGGAG 2726
Db 1947 AGAAAGCTATGGATTCCCTGGATGAAATCCCAAGGAGGCCATCTTCAGAAATGAAAAATG 2006
QY 2727 AAAGGGCCCGAGCCCAAAACCGGTGCGAGGGTGACGGGCCACCCCGGGAGTGGTCTCTAGCACCAAAG 2786
Db 2007 CAA---ACTAAACCCAATGCCAACTTTGTGCAACCTGGAGATCTAGTCTCTAAGCCACACA 2063
QY 2787 CAGGGCCAGCCCTACACATCCGTGTGACACCAAGATCATGAGAACAGCACTGCCACCCCTG 2846
Db 2064 CCTGGCGAGCCACTTCATATAAAGTTACTCCAGACCATGTACAAAACACAGCCACTCTT 2123
QY 2847 GAGATCAAGCCCCACATCTGA-----AGAGTTTCTCTAGTACCACCGTCAATTCCT 2900
Db 2124 GAAATCAAGTCCAAACCAAGAGAGTCTCTCACTCTTACACGAGTACTGCAGTGATACCG 2183
QY 2901 ACCTTAGGCAACACAGAAAACCAAGAAATAACCAATTATTCATCACCCAATGTCTATGCGAA 2960
Db 2184 AACTGTGGCACGCGCAAAAGCAAGGATAACCATCTCTCAAAAACGCTCCATACACCAAGTA 2243
QY 2961 AAGCCCAAAA---GTGCAGATCTTACTCTCGGCCAGAAACGAGCCATGTCCCCTGTCAAG 3017
Db 2244 AAGTCAAAAACCTCTACCGAAGCCTCATGAATTTAGAACAAAGGCATGTCCCCTAATACC 2303
QY 3018 ATTACTACTATTTCCAGAGAGAAAGACCCGGAAGGTGGAAGGAGCGCCTTTGCCGACAGG 3077
Db 2304 ATGSCAACCTTTGCCAGAGACACAGACCCCGAGAGTCTTGTGGTCTTCTTAACCTCCAGAAAG 2363
QY 3078 CCTGCATCCCCCATCCAAATCATGACGGGTGTCAACATCTGCAGTCTCCACTGAAATCGCT 3137
Db 2364 ACAATGT-CCCTATTCAAGTTTGTGGTGTGACTGGTTCAAGTAGCTCTCTGAGCAGGGA 2422
QY 3138 GTCTCTCCTGAATCTCAGGAAGTGCCTATGGGAAGGACTATCCTCAAAGTCAACCCCGAA 3197
Db 2423 CGTCCCCAGAACCAACAGAAATCAGTGCCAAGCATGCGGATATTCAGAGTCTCCCCCAGAC 2482
QY 3198 AAACAAACTGTTCCAGCCCCCGTGGCGGAAGTACAACCTCCAATGCTTAATATCATCACACG 3257
Db 2483 CGGCAGTCATCATGGCAGTTTTCAGCGTTTCAACAGCAATAGCTCAAGTGTGATAACTACT 2542
QY 3258 GAAGACAAATAAAATTCACATTCACCTGGTTCTCAGTTTAAGCGATCT 3305
Db 2543 GAGGATAATAAAATCCACATTCACCTTAGGAAGTCTTACATGCAAGCT 2590
```

RESULT 13
US-10-171-311-47
; Sequence 47, Application US/10171311
; Publication No. US20030087270A1

GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 3025
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-171-311-47

Query Match 12.4%; Score 541.6; DB 15; Length 3025;
Best Local Similarity 53.0%; Pred. No. 3.9e-135;
Matches 1393; Conservative 0; Mismatches 1159; Indels 76; Gaps 8;

```
QY 699 AGTTGAAAAAGCTCCTTGAACAAGAAAAAGTTTACCAAGCCCGCAAGAAAAAGAAAAAC 758
Db 18 AGATTAAAGAGCTAATTGATCAAGAAATCAAGTCTCAGGAGGAGAAAGGACGCAAGAAAAG 77
QY 759 GCTAAGCGGTCAACAAACTTCGAGATGAGCTTGTGAAGCTCAAGTCTTTCGCCCTCATG 818
Db 78 GAGAAAAAGGTCACCCCTGAAAGAGGAGCTGACCAAGCTGAAGTCTTTTGTCTTGATG 137
QY 819 TTGGTGGACGAGAGGCAGATGCACATCGAGCAACTGGGCCCTGCAGAGTCAGAAAAGTCCAG 878
Db 138 GTGGTGGATGAACAGCAAAGGCTGACGGCACAGTCCACCTTTCAAAAGACAGAAAAATCCAA 197
QY 879 GACCTCACTCAGAAAGCTGAGGGAGGAGGAAGAAAACTCAAAAGCGTCACTTACAATCC 938
Db 198 GAGCTGACCACAAATGCAAAAGGAAACACATACCAAACTAGCCCTTGCTGAAGCCAGATT 257
QY 939 AAGCAAGACCCGCCAAGAGCTGTCTCAAGTTAGAGTGGACTTCGAACACAAAGGCCTCAGG 998
Db 258 CAGGAGGAAGAGCAGAAGGCAACCCAGACTAGAGAAGGAACCTGCAAAACGACAGCCACAAAG 317
QY 999 TTTTCCAGGAGCAGAAAGATGAACGCCCAATTTGGCGAATCAAGAAATCTCACAAACCGG 1058
Db 318 TTTTCAACCAAGCAACAGACACAATTTATGGCGAAGCTCACCAATGAGGACAGTCAAAAATCGC 377
QY 1059 CAACCTTCGACTCAAACTGGTTGGCTTATCGCAAAAGGATTGAGGAGTGGAAAGAGACCAAT 1118
Db 378 CAGCTTCAACAAAAGCTGGCAGCACTCAGCCGGCAGATTGATGAGTTAGAAAGAGACAAAC 437
QY 1119 AAAAGCCTTCAGAAAGCAGAGGAGAGCTCCAGGAGCTCAGAGAGAGAGAAAATTTGCCAAAGGG 1178
Db 438 AGGTCTTACGAAAAGCAGAAAGAGAGCTGCAAGATATAAAAGAAAAAATCAGTAAGGGA 497
QY 1179 GAATGTGAAACTCCAGTCTCATGCGCGGAAGTGGAGAGTCTCGCAAGCGCGTGTCTTGAG 1238
Db 498 GAATATGGAACCGCTGCTATCATGCTGAAAGTG----- 530
QY 1239 ATGGAGGGCAAGGATGAAGAGATCAGAAAGACCGAGGCCAGTCCCGGAGAGCTGAAGAG 1298
Db 531 -----GAAGAGCTCATAAAAAATGGAGGAGCAGTGCAGAGAGATCTCAATAAG 575
```

QY 1299 AAGCTCCAAGAGGAAGAACACACAGCAAGGAACCTTAGACTAGAAAGTGGAGAAAGCTGCAG 1358
Db 576 AGGCTTGAAAGGGAGACGTTACAGAGTAAAGACTTTAACTAGAGGTTGAAAACTCAGT 635
QY 1359 AAGAGGATGCTGAGCTGGAGAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGC 1418
Db 636 AAAAGAAATTATGGCTCTGGAAAAGTTAGAAGACGCTTTCAACAAAAAGCAAAACAAGAATGC 695
QY 1419 ACCCAGCTCCATCTGAACCTGGAGAAAGGAGAAAGAACCTTAACCAAAAGACCTGCTGAACGAG 1478
Db 696 TACTCTCTGAATGCAATTTAGAAAAAGAAAGGATGACCACAAAGCAGTTGTCTCAAGAA 755
QY 1479 CTGGAGGTGGTCAAGAGTCAAGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAG 1538
Db 756 CTGGAGAGTTTAAAAGTAAGGATCAAAGAGCTAGAAGCCATTGAAAGTCGGCTAGAAAAG 815
QY 1539 GCCGAGTTAAGCCTCAAAGATGACCTTACAAAAGCTGAAGTCCCTTCACTGTGATGCTGGTG 1598
Db 816 ACAGAATTCACTCTAAAGAGGATTTAACTAACTGAACCATTAACCTGTGATGTTGTA 875
QY 1599 GATGAGAGGAAAAATATGATGGAGAAAAATAAGCAAGAGAGAGGAAAGTGGATGGTTG 1658
Db 876 GATGAACGGAAAAACAATGAGTGAAAAAATTAAGAAAACTGAAGATAAAATTACAGCTGCT 935
QY 1659 AATAAAAACTTTAAGGTGAGCAGGGAAAAAGTCAATGATGTGACGGAAAAAGCTAATCGAG 1718
Db 936 TCTTCTCAGCTTCAAGTGGAGCAAAAATAAAGTAACAACAGTTACTGAGAAGTTAATTGAG 995
QY 1719 GAAAGCAAGAAAGCTTTTAAAAACTCAAATCTGAAATGAGGAAAAAGGAGTACAGTGTGACA 1778
Db 996 GAAACTAAAAGGGCGCTCAAGTCCAAAACCGATGTAGAAGAAAAGATGTACAGCGTAAAC 1055
QY 1779 AAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGCAAGAAAGAAAGTCCCTGTGAAC TG 1838
Db 1056 AAGGAGAGAGATGATTTAAAAAACAAATTGAAAGCGGAAGAGAGAAAGGAATGATCTC 1115
QY 1839 AGCTGCAGTGTAGACTTACTAAAGAAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGGGAA 1898
Db 1116 CTGTCAAGAGTTAATATGTTGAAAAATAGGCTTCAATCATTTGGAAGCAATTGAGAAAGAT 1175
QY 1899 ATAAACCGAGGTAGTGGTCAAGGGGTCTGAGTTCAAC-----TGCCCGGAA 1946
Db 1176 TTCTTAAAAACAATTAATCAAGACTCTGGGAAATCCACAACAGCATTACACCAAGAA 1235
QY 1947 GACAATAAGATCAGAGAACTAAACGCTTGAAATCGAGAGACTGAAGAAACCGCTCCAGCAG 2006
Db 1236 AACAATAAGATTAAAGAGCTCTCTCAAGAAAGTGGAAAGACTGAACTGAAGCTTAAAGGAC 1295
QY 2007 TTGGAGGTGGTGGAGGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAGCAG 2066
Db 1296 ATGAAAGCCATTGAGGATGACCTCATGAAAACAGAAAGATGAATATGAGACTCTAGAACGA 1355
QY 2067 AAGTTCAGAACCGCAGGATAAGGCCAAACCTTCTCTCCAGCAGCTCGAGGAAATCAA 2126
Db 1356 AGGTATGCTAATGAACGAGACAAAGCTCAATTTTATCTAAAGAGCTAGAACATGTTAAA 1415
QY 2127 CACCAAATGGCCAGCACAAAGCCATAGAGAAAGGGAGGCGGTGAGCCAGGAAGCCGAA 2186
Db 1416 ATGGAACCTTGCTAAGTACAAGTTAGCAGAAAGAGACAGAGACC---AGCCATGAACAATGG 1472
QY 2187 CTGCGACACAGGTTTCGGCTGGAGGAGGCTTAAAGTCGTGATTTACAGGCCGAGGTGCAG 2246
Db 1473 CTTTTCAAAAGGCTTCAAGAAGAAGAAAGCTAAGTCAGGGCACCTCTCAAGAGAAGTGGAT 1532
QY 2247 GCTCTCAAGGAGAGATCCACGAGCTGATGAACCAAGGAAGACCAGCTGTCTCAGCTCCAA 2306
Db 1533 GCATTAAAGAGAAAAATTTCATGAATACATGGCCTGAAGACCTAATATGTACCTCCAG 1592
QY 2307 GTCGACTATTCCGTCTTCAGCAAAGATTTTATGGAAGAAGAAAATAAGAAACAAGAACATG 2366
Db 1593 GGAGATCACTCAGTCTGCAAAAAAATACTAAATCAACAAGAAAAACAGGAACAGAGATTTA 1652
QY 2367 GGGAGGGAGGTCCTCAATCTGTACCAAGGAGCTAGAGCTTTCCAAGCGCTACAGCCGAGCT 2426

Db 1653 GGAAGAGAGATTGAAAAACCTCACTAAGGAGTTAGAGAGGTACCGGCATTTTCAGTAAGAGC 1712
QY 2427 CTCAGGCCGAGTGGGAAACGGCCGAAGGATGGTGGACGTGCCTGTGGCCTCCACTGGGGTG 2486
Db 1713 CTCAGGCCTAGTCTCAATGGAAGAAAGATTTCCGATCCTCAAGTATTTTCTTAAAGAAAGTT 1772
QY 2487 CAGACCGAGGCGGTGTCGGGGGATGCTGCGGAGGAGGAGACCCCGCTGTGTTTCATTCGC 2546
Db 1773 CAGACAGAAGCAGTAGACAATGAACCACTGATTACAAGAGCCCTCATTTCTCTGGAACGT 1832
QY 2547 AAATCCTTCCAGGAGGAAAAATCACATCATGAGTAATCTTTCGACAGGTAGGCCTGAAGAAA 2606
Db 1833 GCAGTCATCAATGGTCAGTTATATGAGGAGAGTGAGAAATCAAGACGAGGACCCTAATGA- 1891
QY 2607 CCCATGGAACGGTCCCTCGGTCTCTGACAGGTATCCCCAGCAGCGAATGAGTCCACATG 2666
Db 1892 -----TGAGGGATCTGTGCTGCTCTTCAAATGCAGCCAGTCTACTCCATGTCTGTTAAC 1946
QY 2667 AGGAAGTCTTGGATTCTTGGATGAGAAAAAGAGAAAAACGGTCTCTTCCACTCCGACAGGAG 2726
Db 1947 AGAAAGCTATGGATTCCCTGGATGAAATCCAAGGAGGCCATCTTTCAGAAATGGAATAATG 2006
QY 2727 AAAGGGCCAGGCCAAACAGGGTGCAGGGCACCCCGGGGAGCTGGTCTCTAGCACCAAAAG 2786
Db 2007 CAA---ACTAAACCCCAATGCCAACTTTGTGCAACTTGAGATCTAGTCTTAAGCCACACA 2063
QY 2787 CAGGGCCAGCCCCTACACATCCGTGTGACACCAGATCATGAGAACAGCACTGCCACCTG 2846
Db 2064 CTTGGCAGCCACTTCATATAAAGGTTACTCCAGACCATGTACAAAACACAGCCACTCTT 2123
QY 2847 GAGATCACAAGCCCCACATCTGA-----AGAGTTTTTCTTAGTACCACCGTCAATTCCT 2900
Db 2124 GAAATCACAAGTCCAAACACAGAGAGTCTCTCACTCTTACACGAGTACTGCAGTGATACCG 2183
QY 2901 ACCTTAGGCAACCAGAAACCAAGAAATAACCATTATTCATCACCCCAATGTCTATGTCGCA 2960
Db 2184 AACTGTGGCAGCGCAAGCAAGGATAACCATCTCCAAAACGCCCTCCATAACACCAGTA 2243
QY 2961 AAGCCCCAAA---GTGCAGATCCTACTCTCGGCCCAGAACGAGCCATGTCCCTGTTCACG 3017
Db 2244 AAGTCCAAAACCTCTACCGAAGACCTCATGAATTTAGAACAAAGGCATGTCCCAATTACC 2303
QY 3018 ATTACTACTATTTCCAGAGAGAAAGAGCCCCGGAAGGTGGAAGGAGCGCCTTTGCCGACAGG 3077
Db 2304 ATGGCAACCTTTGCCAGAGCACAGACCCAGAGTCTTGTGTTCTCTAACTCCAGAAAGG 2363
QY 3078 CTTGCATCCCCATCCAAATCATGACGGTGTCAACATCTGCAGTCTCCACTGAAATCGCT 3137
Db 2364 ACAATGT-CCCTAATTCAGGTTTGGCTGTGACTGGTTGAGTCTCTCTGAGCAGGGA 2422
QY 3138 GTCTCTCTGAAATCTCAGGAAGTGCCTATGGAAGGAGTATCCTCAAAGTCACCCCGGAA 3197
Db 2423 CGTCCCCAGAACCAACAGAAATCAGTGCCTCAAGCATGCGATATTCAGAGTCTCCCCAGAC 2482
QY 3198 AAACAAACTGTTCCAGCCCCCGTGCAGGAGTACAACTCCAATGCTAATATCATCACCACG 3257
Db 2483 CGGCAGTCATCATGGCAGTTTCAGCGGTTCAAACAGCAATAGCTCAAGTGTGATAACTACT 2542
QY 3258 GAAGACAAATAAAATTCACATTCACCTGGGTTCTCAGTTTAAAGCGATCT 3305
Db 2543 GAGGATAATAAAATCCACATTCACTTAGGAAGTCTTACATGCAAGCT 2590

RESULT 14

US-10-309-851-23

; Sequence 23, Application US/10309851

; Publication No. US20030108554A1

; GENERAL INFORMATION:

; APPLICANT: Saus, Juan

; APPLICANT: Revert-Ros, Francisco

; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; INTERACT WITH GOODPASTURE ANTIGEN BINDING PROTEIN

; FILE REFERENCE: 98,723-F-US
; CURRENT APPLICATION NUMBER: US/10/309,851
; CURRENT FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 2355
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2355)
; OTHER INFORMATION:
;
US-10-309-851-23

Query Match 11.4%; Score 496.6; DB 15; Length 2355;
Best Local Similarity 53.8%; Pred. No. 5.2e-123;
Matches 1174; Conservative 0; Mismatches 974; Indels 33; Gaps 6;

QY 1146 CTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAAACTTCCAGTCTCATGGCG 1205
Db 1 CTGCAAGATATAAAGAAAAAATCAGTAAGGGAGATATGGAAACGCTGGTATCATGGCT 60
QY 1206 GAAGTGGAGAGTCTGCGCAAGCGGTGCTTGAGATGGAGGCAAGGATGAAGAGATCAGC 1265
Db 61 GAAGTGGAGAGCTCAGGAAACGTGTGCTAGATATGGAGGGAAAGATGAAGAGCTCATA 120
QY 1266 AAGACCGAGGCCAGTCCGGGAGCTGAAGAAAGCTCCAAGAGGAAGAACACCACAGC 1325
Db 121 AAAATGGAGGAGCAGTGCAGAGATCTCAATAAGAGGCTTGAAGGGAGAGCGTTACAGAGT 180
QY 1326 AAGGAACTTAGACTAGAACTGGAGAAAGCTTCAGAAAGAGGATGTCTGAGCTGGAGAAGCTG 1385
Db 181 AAAGACTTTAAACTAGAGGTGAAAAACTCAGTAAAGAAATATATGGCTCTGGAAAAAGTTA 240
QY 1386 GAGGAAGCGTTCCAGCCGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAG 1445
Db 241 GAAGACGCTTCAACAAAGCAACAAAGAAATGCTACTCTCTGAAATGCAATTTAGAAAAA 300
QY 1446 GAGAAGAACTTAACCAAGACCTGCTGAACAGAGCTGGAGGTGGTCAAGAGTCGAGTTAAA 1505
Db 301 GAAAGGATGACCAACAAAGCAGTGTCTCAAGAACTGGAGAGTTTAAAGTAAGGATCAAA 360
QY 1506 GAACTCGAATGCTCCGAGAGTAGACTGGAGAAAGCCGAGTTAAGCCTCAAAGATGACCTT 1565
Db 361 GAGCTAGAACCCATTGAAAGTCGGCTAGAAAGACAGAAATCACTCTAAAAGAGGATTTA 420
QY 1566 ACAAGCTGAAAGTCTTCACTGTGATGCTGTGGATGAGAGGAAAAATATGATGGAGAAA 1625
Db 421 ACTAACTGAAAAACATTAACTGTGATGTTGTAGATGAACGGAACAAATGAGTGAAAAA 480
QY 1626 ATAAAGCAAGAAGAGAGGAAAGTGGATGGGTTGAATAAAAACTTTAAGGTGGAGCAGGGA 1685
Db 481 TTAAGAAAAACTGAAGATAAATTACAAGCTGCTTCTTCTCAGCTTCAAGTGGAGCAAAAT 540
QY 1686 AAAGTCTATGGATGTGACCGGAAAAAGCTAATCGAGGAAAGCAAGAAAGCTTTTAAAACTCAA 1745
Db 541 AAAGTAAACAAGTTACTGAGAAATTAAATTGAGGAAACTTAAAGGGCGCTCAAGTCCAAA 600
QY 1746 TCTGAAATGGAGGAAAAAGGAGTACAGTCTGACAAAGGAGAGGAGATGAGCTGATGGGTAAA 1805
Db 601 ACCGATGTAGAGAAAGAGATGTACAGCGTAAACCAAGGAGAGAGATGATTTAAAAAAACAAA 660
QY 1806 CTGAGGAGCGGAAGAAAGGTCCTGTGAACCTGAGCTGCGAGTGTAGACTTACTTAAAGAAAG 1865
Db 661 TTGAAAGCGGAAGAAAGAGAAAGGAAATGATCTCTCTGTCAAGAGTTAATATGTTGAAAAAT 720
QY 1866 CGGCTTGATGGCATAGAGGAGGTAGAAAGGGAATAAAACCGAGGTAGGTCGTGCAAGGGG 1925
Db 721 AGGCTTCAATCATTTGGAAGCAATTGAGAAAGATTTCTTAAAAAACAAATTTAAATCAAGAC 780
QY 1926 TCTGAGTTTCACC-----TGCCCGGAAGACAATAAGATCAGAGAACTTAACGCTT 1973

Db 781 TCTGGAAAAATCCACAACAGCATTACACCAAGAAAAACAATAAGATTAAAGAGCTCTCTCAA 840
QY 1974 GAAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTTGAGAGTGGTGGAGGGGACTTGTATG 2033
Db 841 GAAGTGGAAAGACTGAAGAACTGAAGCTAAAGGACATGAAGCCATTTGAGGATGACCTCATG 900
QY 2034 AAGACCGAGGACGAATATGACCAGTTGAGAGCAGAAAGTTTCAGAAACCGAGCAGGATAAGGCA 2093
Db 901 AAAACAGAAAGATGAATATGAGACTCTAGAAACGAAGGTATGCTAATGAACGAGACAAAGCT 960
QY 2094 AACTTCTCTCCCAGCAGCTCGAGGAAATCAAACACCAAATGGCCAAAGCAAAAGCCATA 2153
Db 961 CAATTTTATCTAAAGAGCTAGAACATGTTAAAAATGGAACTTGCTAAGTACAAGTTAGCA 1020
QY 2154 GAGAAAGGGGAGGCCGTGAGCCAGGAAGCCGAACTGCGACACAGGTTTCGGCTGGAGGAG 2213
Db 1021 GAAAAGACAGAGACC---AGCCATGAACAATGGCTTTTCAAAAAGGCTTCAAGAAAGAGAA 1077
QY 2214 GCTAAAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAGGAGAGAAAGATCCACGAGCTG 2273
Db 1078 GCTAAGTCAGGGCACCTCTCAAGAGAAAGTGGATGCAATTAAGAAGAGAAAAATTCATGAATAC 1137
QY 2274 ATGAAACAAGGAAGACCAGCTGTCTCAGTCCAAAGTCGAATTCGATTCGTTCCGTCAGCAAGA 2333
Db 1138 ATGGCAACTGAAGACCTAATATGTCACTCCAGGAGATCACTCAGTCTCTGCAAAAAAAA 1197
QY 2334 TTTATGGAAGAAAGAACTAAGAAACAAGAACATGGGGAGGAGGTCTCAATCTGACCAAG 2393
Db 1198 CTAATCAACAAGAAACACAGGAACAGAGATTTAGGAAGAGAGATTTGAAAACTCCTCACTAAG 1257
QY 2394 GAGCTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTCTCAGGACGGCCGNAAGG 2453
Db 1258 GAGTTAGAGAGGTACCCGCCATTTTCAGTAAGAGCCTCAGGCCTAGTCTCAATGGAAGAAAGA 1317
QY 2454 ATGGTGAACGTGCTGTGGCCTCCACTGGGGTGCAGACCGAGGCGGTGTGCGGGGATGCT 2513
Db 1318 ATTTCCGATCCTCAAGTATTTTCTAAAGAAAGTTTCAGACAGAAAGCAGTAGACAATGAACCA 1377
QY 2514 GCGGAGGAGGAGACCCCGCTGTGTTTCATTCGCAAAATCTTCCAGGAGGAGGAAAAATCACATC 2573
Db 1378 CCTGATTACAAGAGCCTCATTCCTCTGGAACGTGCAGTCAATCAATGGTCACTATATAG 1437
QY 2574 ATGAGTAAATCTTCGACAGGTAGCCCTGAAGAAACCCATGGAACGGTCTCTCGGTCTCTCGAC 2633
Db 1438 GAGAGTGAGAAATCAAGACGAGGACCCCTAATGAT-----GAGGGATCTGTGCTGCTCTTC 1491
QY 2634 AGGTATCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTGATTCCTTTGGATGAGA 2693
Db 1492 AAATGCAGCCAGTCTACTCCATGTCTCTGTAAACAGAAAGCTATGGATTCCCTGGATGAAA 1551
QY 2694 AAAAGAGAAAAACGGTCTTCCACTCCGACAGGAGAAAGGGCCAGGCCCAACCCAGGGTGCA 2753
Db 1552 TCCAAGGAGGGCCATCTTCAGAAATGAAAAAATGCAAA---CTAAACCCCAATGCCAACTTT 1608
QY 2754 GGGCACCCCGGGGAGCTGGTCTTAGCACCAAGCAGGGCCAGCCCCCTACATCCGTGTG 2813
Db 1609 GTGCAACCTGGAGATCTAGTCTTAAGCCACACACCTGGGCAGCCACTTCATATAAAGGTT 1668
QY 2814 ACACCAGATCATGAGAACAGCACTGCCACCTGGAGATCACAGCCCCCACATCTGAAG-- 2871
Db 1669 ACTCCAGACCATGTACAAAACACAGCCACTCTTGAAATCACAAGTCCAACCCACAGAGAGT 1728
QY 2872 ---AGTTTTTCTCTAGTACCACCGTCACTTCTACCTTAGGCAACCCAGAAACCAAGAATA 2927
Db 1729 CCTCACTCTTACACGAGTACTGTCAGTGTATACCGAACTGTGGCAGCCCAAGCAAAAGGATA 1788
QY 2928 ACCATTATTCATCACCCAAATGTTCATGTGCGAAAAGCCCAAAA---GTGCAGATCCTACT 2984
Db 1789 ACCATCTCCAAAACGGCTCCATAACACCGTAAGTCCAAAACCTCTACCCGAAGACCTC 1848
QY 2985 CTCGGCCCCAGAACGAGCCATGTCCCTGTCTACGATTACTACTATTTCAGAGAGAGAGAGC 3044
Db 1849 ATGAATTTAGAAACAAGGCGATGTCCCAATTTACCATGGCAACCTTTTGCCAGAGCAGACCC 1908

THIS PAGE BLANK (6847)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 7, 2004, 20:50:29 ; Search time 15998 Seconds
(without alignments)
8145.928 Million cell updates/sec

Title: US-10-788-793-1
Perfect score: 4364
Sequence: 1 ccactgggtttctcaaggga.....aaaaaaaaaaaaaaaaaaaa 4364

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: em_gss_hum:*
 - 18: em_gss_inv:*
 - 19: em_gss_pln:*
 - 20: em_gss_vrt:*
 - 21: em_gss_fun:*
 - 22: em_gss_mam:*
 - 23: em_gss_mus:*
 - 24: em_gss_pro:*
 - 25: em_gss_rod:*
 - 26: em_gss_phg:*
 - 27: em_gss_vrl:*
 - 28: gb_gss1:*
 - 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1412.4	32.4	2079	11 AK017709	AK017709 Mus muscu
2	687.2	15.7	1080	11 BC029425	BC029425 Homo sapi
3	627.2	14.4	826	10 BF585363	BF585363 602102141
4	621.6	14.2	716	14 CF736504	CF736504 UI-M-HD0-

5	620.8	14.2	3240	11	AK019472	AK019472 Mus muscu
6	615.4	14.1	618	9	AI176515	AI176515 EST220101
7	615.4	14.1	777	11	BC030632	BC030632 Mus muscu
8	606	13.9	606	14	CB581318	CB581318 AMGNNUC:N
9	581.6	13.3	844	9	AU130457	AU130457 AU130457
10	576.2	13.2	583	10	BF562820	BF562820 UI-R-BJ0p
11	573.2	13.1	582	10	BE095605	BE095605 UI-R-B01-
12	561.2	12.9	785	12	BG213104	BG213104 RST32708
13	561	12.9	586	10	BF398646	BF398646 UI-R-BS2-
14	553	12.7	553	10	AW525176	AW525176 UI-R-BJ0p
15	540	12.4	540	14	CB612875	CB612875 AMGNNUC:N
16	536.6	12.3	600	12	BG800198	BG800198 2113-92 M
17	534.2	12.2	700	14	CF738763	CF738763 UI-M-HD0-
18	532.8	12.2	671	14	CF727145	CF727145 UI-M-HB0-
19	524.4	12.0	691	13	BX670884	BX670884 BX670884
20	519.4	11.9	600	13	BU920550	BU920550 6061-68 M
21	501	11.5	669	13	BX507241	BX507241 DKFZp7791
22	496.4	11.4	518	10	BF411781	BF411781 UI-R-BT1-
23	495.2	11.3	503	10	BE097165	BE097165 UI-R-B01-
24	490	11.2	600	14	CA530281	CA530281 9033-39 M
25	487	11.2	535	12	BM114762	BM114762 L0809H10-
26	460.2	10.5	845	13	BU339886	BU339886 603514850
27	443.8	10.2	505	9	AI390999	AI390999 mc04b10.Y
28	431.8	9.9	611	14	CF789133	CF789133 866381 MA
29	431	9.9	479	10	BF412113	BF412113 UI-R-BT1-
30	430.8	9.9	466	13	BQ555036	BQ555036 H4031G04-
31	423	9.7	629	14	CD685106	CD685106 EST1626 h
32	423	9.7	755	13	BU448785	BU448785 603213038
33	417.8	9.6	435	13	BQ194329	BQ194329 UI-R-CN1-
34	414.2	9.5	529	13	EX109059	EX109059 BX109059
35	414	9.5	414	10	AW532783	AW532783 UI-R-BU0-
36	411.6	9.4	610	12	BM440306	BM440306 pgrln.pk0
37	406.2	9.3	459	13	BY007025	BY007025 BY007025
38	405.4	9.3	645	9	AV241894	AV241894 AV241894
39	401	9.2	791	13	BU401987	BU401987 603483217
40	400	9.2	826	13	EX736046	EX736046 BX736046
41	399.6	9.2	1005	14	W29517	W29517 mc03c05.r1
42	399.2	9.1	603	9	AL596735	AL596735 DKFZp451B
43	399	9.1	535	9	AL044677	AL044677 DKFZp434C
44	398	9.1	528	9	AA822322	AA822322 vw36f04.r
45	396	9.1	621	13	BU219496	BU219496 603755856

ALIGNMENTS

RESULT 1

AK017709

LOCUS

DEFINITION

AK017709 2079 bp mRNA linear HTC 20-SEP-2003
Mus musculus 8 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:5730485H21 product:hypothetical
Tropomyosin/Glutamic acid-rich region containing protein, full
insert sequence.

ACCESSION

AK017709

VERSION

AK017709.1 GI:12857088

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE

1

AUTHORS

Carninci,P. and Hayashizaki,Y.

TITLE

Normalization and subtraction of cap-trapper-selected cDNAs to

JOURNAL

Meth. Enzymol. 303, 19-44 (1999)

MEDLINE

99279253

PUBMED

10349636

REFERENCE

2

AUTHORS

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE

Normalization and subtraction of cap-trapper-selected cDNAs to

JOURNAL

Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE

20499374


```
QY 783 GATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCATGTTGGTGGACGAGAGGCAGATGCAC 842
Db |||||||
QY 843 ATCGAGCAACTGGGCCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAGCTGAGGGAG 902
Db |||||||
QY 903 GAGGAAGAAAACTCAAAGCGGTCACTTACAAATCCAAAGAACCGCCAGAGCTGCTC 962
Db |||||||
QY 963 AAGTTAGAAAGTGGACTTCGAACACAAAGGCCTCGAGGTTTCCAGGAGCACGAAGAGATG 1022
Db |||||||
QY 1023 AACGCCAAATTGGCGAATCAAGAATCTCAACACCGGCAACTTCGACTCAAACTGGTTGGC 1082
Db |||||||
QY 1083 TTATCGCAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAGGCAGAGGAA 1142
Db |||||||
QY 1143 GAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAACCTCCAGTCTCATG 1202
Db |||||||
QY 1203 GCGGAAGTGGAGAGTCTGCGCAAGCGCGTCTTGAGATGGAGGGCAAGGATGAAGAGATC 1262
Db |||||||
QY 1263 ACGAAGACCGAGGCCCTAGTCCCGGGAGCTGAAGAAAGAGCTCCAAGAGGGAAGAACCCAC 1322
Db |||||||
QY 1323 AGCAAGGAACTTAGACTAGAAGTGGAGAAGCTGCAGAAGAGGATGTCTGAGCTGGAGAAG 1382
Db |||||||
QY 1383 CTGGAGGAAGCGTTACCGCGAGTAAGTCGGAATGCACCCAGCTCCATCTGAAGCTGGAG 1442
Db |||||||
QY 1443 AAGGAGAAGAACCTTAACCAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCCAGTT 1502
Db |||||||
QY 1503 AAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAGAGCCGAGTTAAGCCTCAAAGATGAC 1562
Db |||||||
QY 1563 CTTACAAAGCTGAAGTCTTCACTGTGATGCTGTGGATGAGAGGAAAAATATGATGG 1620
Db |||||||
QY 2022 CTCACGAAGCTGAAGTCTTCACTGTGATGCTGTGGTGGACGAGAGGAAAAACATGATGG 2079
```

```
RESULT 2
BC029425
LOCUS BC029425
DEFINITION Homo sapiens, Similar to KIAA1275 protein, clone IMAGE:4616553, mRNA.
ACCESSION BC029425
VERSION BC029425.1
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1080)
AUTHORS Strausberg,R.
TITLE Direct Submission
```

JOURNAL

REMARK
COMMENT

Submitted (01-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: CLONTECH

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) mcd@paxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 41 Row: g Column: 14

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis

This clone has the following problem: frame shifted.

FEATURES

source

Location/Qualifiers

1..1080

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4616553"

/tissue_type="Lung"

/clone_lib="NIH MGC_77"

/lab_host="DH10B"

/note="Vector: pDNR-LIB"

ORIGIN

Query Match 15.7%; Score 687.2; DB 11; Length 1080;

Best Local Similarity 83.6%; Pred. No. 4.4e-93;

Matches 779; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

```
QY 699 AGGTTGAAAAAGCTCCTTGAACAAGAAAAAGCTTACCAAGCCCGCAAGAAAAAGGAAAAAC 758
Db |||||||
QY 147 AGGTTAAAAAAGCTCCTTGAACAAGAAAAAGCTTATCAAGCCCGCAAGAAAAAGGAAAAAT 206
Db |||||||
QY 759 GCTAAGCGGCTCAACAAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCATG 818
Db |||||||
QY 207 GCTAAACGACTCAATAAACAATAAGAGATGAGCTTGTAAACTCAAATCCTTTGCACCTATG 266
Db |||||||
QY 819 TTGGTGGACGAGAGCGAGATGCACATCGAGCAACTGGGCTGCAGAGTCAGAAAGTCCAG 878
Db |||||||
QY 267 CTGGTGGATGAAGACAAATGCACATTGAACAACTTGGCTGCAAAGCCAGAAAGTACAG 326
Db |||||||
QY 879 GACCTCACTCAGAAAGCTGAGGGAGGAGGAAGAAAACTCAAAGCGGTCACTTACAAATCC 938
Db |||||||
QY 327 GATCTTACTCAGAAAGCTGAGGGAAGAAAGAGAGAGCTCAAAGCCATTACTTCCAAATCC 386
Db |||||||
QY 939 AAGGAAGACCGCCAGAAAGCTGCTCAAGTTAGAGTGGACTTCGAACACAAAGGCTCGAGG 998
Db |||||||
QY 387 AAAGAAGACAGACAGAAAAATTGCTCAAGTTAGAGTGGACTTTGAACACAAAGGCTTCGAGG 446
Db |||||||
QY 999 TTTTCCCAGGAGCACCAAGAGATGAACCGCAAAATTTGGCGAATCAAGAAATCTCACAACCGG 1058
Db |||||||
QY 447 TTTTCTCAAGAGCATGAAGAGATGAACGCTAAAAGTGGCTAATCAAGAGTCTCACAATAGG 506
Db |||||||
QY 1059 CAACTTCGACTCAAACCTGGTTGGCTTATCGCAAAGGATTCGAGGAGCTGGAAGAGACCAAT 1118
Db |||||||
QY 507 CAACTTAGACTCAAGCTGGTTGGCTTAACCCAAAGAAATCGAGGAGCTAGAAAGAGACCAAC 566
Db |||||||
QY 1119 AAAAGCCTTCAGAAAGCGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAAATGCCCCAAGGG 1178
Db |||||||
QY 567 AAAAATCTGCAGAAAGCGCAGAGGAAGAACTTCAAGAAATTAAGAGATATAAAATTGCCAAAGGA 626
Db |||||||
QY 1179 GAATGTGGAACCTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTGTGAG 1238
```

```

|||||
Db 627 GAATGTGGAACCTCTAGCCTCATGGCAGAAGTGGAATAATCTTCGAAAGCGTGTGCTTGAA 686
|||||
QY 1239 ATGGAGGGCAAGGATGAAGAGAGATCACGAAGACCGAGGCCAGTGCCGGGAGCTGAAGAAG 1298
|||||
Db 687 ATGGAAGGTAAAGATGAGGAGATCACTAAAACTGAAATCCAGTGTAGGGAATTGAGGAAG 746
|||||
QY 1299 AAGCTCAAGAGGAAGAAACACACAGCAAGGAACCTTAGACTAGAAGTGGAGAAGCTGCAG 1358
|||||
Db 747 AAGCTGCAAGAGGAAGAAACACCATAGTAAGGAGCTCAGACTTGAAGTTGAGAAGCTACAG 806
|||||
QY 1359 AAGAGGATGCTGTGAGCTGGAGAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAAGTCGGAATGC 1418
|||||
Db 807 AAGAGAAATGCTCTGAACCTAGAGAAATTTGGAAGAAAGCAATTTAGCAAGAGTAAATCTGAGTGC 866
|||||
QY 1419 ACCCAGCTCCATCTGAACCTGGAGAAAGGAAGAACCTTAACCAAGACCTGCTGAACGAG 1478
|||||
Db 867 ACCCAGCTACATTTAAATCTGGAGAAAGAAAGAAACCTTAACCAAGACCTGCTAAATGAA 926
|||||
QY 1479 CTGGAGGTGGTCAAGAGTCGAGTTAAAGRAACTCGAATGCTCCGAGAGTAGACTGGAGAAG 1538
|||||
Db 927 TTGGAGGTGGTCAAGAGTCGAGTTAAAGAAATTTGGAATGTTCTGAAAGTAGAATTTGAAAAAG 986
|||||
QY 1539 GCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAAGTCCTTCACTGTGATGCTGGTG 1598
|||||
Db 987 GCTGAATTAAGCCTAAAGATGATCTTTACCAAGTTGAAGTCATTTACCCTGATGCTGGIT 1046
|||||
QY 1599 GATGAGAGGAAAAATATGATGGAGAAAAATAAA 1630
|||||
Db 1047 GATGAAAAAATAAATAAATAAATAAATAAATAAATAA 1078
|||||
```

```

RESULT 3
BF585363 826 bp mRNA linear EST 12-DEC-2000
LOCUS 602102141F2 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4225066 5',
DEFINITION mRNA sequence.
```

```

ACCESSION BF585363
VERSION BF585363.1 GI:11659081
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 826)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparedation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9816 row: b column: 11
High quality sequence stop: 766.
Location/Qualifiers
1. .826
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4225066"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Co24"
/note="Organ: cGlon; Vector: pCMV-sport6; Site_1: NotI;
Site_2: Sali; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
```

FEATURES

source

```

RESULT 4
CF736504 716 bp mRNA linear EST 10-OCT-2003
LOCUS UI-M-HD0-ckp-m-10-0-UI.r1 NIH_BMAP_HD0 Mus musculus cDNA clone
DEFINITION IMAGE:30612297 5', mRNA sequence.
ACCESSION CF736504
VERSION CF736504.1 GI:37632840
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 716)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
```

ORIGIN

```

Query Match 14.4%; Score 627.2; DB 10; Length 826;
Best Local Similarity 91.8%; Pred. No. 4.5e-84;
Matches 696; Conservative 0; Mismatches 58; Indels 4; Gaps 3;
QY 1980 GAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTGGAGGGGACTTGATGAAGACC 2039
Db 1 GAGAGACTGAAGAAACGGCTCCAGCAGCTGGAGGTGGAGGGGACTTGATGAAGACC 60
QY 2040 GAGGACGAATATGACCAGTTGGAGCAGAAGTTTCAGAACCGAGCAGGATAAGGCCAAACTTC 2099
Db 61 GAGGACGAATA CGACCAGTTGGAGCAGAATTCAGAACCGAGCAGGATAAGGCCAAATTC 120
QY 2100 CTCTCCAGCAGCTCGAGGAAATCAAAACACCAAAATGGCCAAAGCACAAAGCCATAGAGAAA 2159
Db 121 CTCTCCAGCAGCTCGAGGAAATTAAGCACCAAAATGGCCAAAGAACAAAGCCATAGAGAAA 180
QY 2160 GGGGAGGCCGTGAGCCAGGAAAGCCGAACACACAGTTTCGGCTGGAGGAGGCTAAA 2219
Db 181 GGGGAGGCTGTGAGCCAGGAAAGCCGAGCTGCGACACAGGTTCCGGATGGAGGAGGCTAAA 240
QY 2220 AGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAGGAGAAAGATCCACGAGCTGATGAAC 2279
Db 241 AGTCGTGACTTACAGGCTGAGGTGCAGGCTCTTAAAGAGAAAGATCCACGAGCTGATGAAC 300
QY 2280 AAGGAAGACCAGCTGTCTCAGCTCCAAAGTCGACTATTCCGTCCTTCAGCAAAGATTATG 2339
Db 301 AAGGAAGATCAGCTCTCTCAGCTCCAGTCCAGTCCAGTCTTCCGTCCTTCAGCAAAGGTTATG 360
QY 2340 GAAGAAGAAACTAAGAACAAAGAACATGGGAGGGAGGTCTCTCAATCTGACCAAGGAGCTA 2399
Db 361 GAAGAAGAAACTAAGAACAAAGAACATGGGAGGGAGGTCTCTCAACCTTGACCAAGAGCTG 420
QY 2400 GAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGGCCGAGTGGGAAACGGCCGAAAGGATGGTG 2459
Db 421 GAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGGCCGAGTGGGAAATGGCCGAAAGGATGGTG 480
QY 2460 GACGTGCTGTGGCTCCACTGGGGTGACAGCCGAGCGGTGTGCGGGGATGCTGCGGAG 2519
Db 481 GACGTACCCGTGGCTCCACTGGGGTGACAGCCGAGCGGTGTGCGGGGATGCTGCGGAG 540
QY 2520 GAGGAGACCCCGGCTGTTCATTCGCAATCCTTCCAGGAGGAAATCA-CATCATGAG 2578
Db 541 GAGGAGACCCCGGCTGTTCATCCGCAATCCTTCCAGGAGGAAATCA-CATCATGAG 600
QY 2579 TAATCTTCGACAGGTAGGCCTG-AAGAAACCCATGGAACGGTCTCGGTCTCGACAGGT 2637
Db 601 TAATCTTCGACAGGTAGGCCTGAAAGAAACCCATGGAACGGTCTCGGTCTCGACAGGT 660
QY 2638 ATCCCCCAGCAGCAATGAGTCCCATGAGGAA--GTCTTGATTCTTGGATGAGAAA 2695
Db 661 TATCCCCCAGCCGCAATGAATCACCATGAGGAAAGTTCTTTGGATTCTTGGATGAGAAA 720
QY 2696 AAGAGAAAAACGGTCTCTCCACTCCGACGAGGAGAAAAAGGC 2733
Db 721 AACGAGAAAAACGGGACCTCAGTTCCAGGAGGAGAGGGCC 758
```

RESULT 4

CF736504

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
On Dec 6, 2002 this sequence version replaced gi:12859698.
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCGAGTTAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised longer than 7 kb was selected before cloning. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

FEATURES

source
1..3240
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:4631422005"
/db_xref="MGI:1912759"
/db_xref="taxon:10090"
/clone="4631422005"
/tissue_type="skin"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stages="0 day neonate"
misc_feature
1..3240
/note="DOC1 homolog [Homo sapiens] (SPTR|Q13597, evidence: FASTY, 88.3%ID, 100%length, match=2299)"

ORIGIN

Query Match 14.2%; Score 620.8; DB 11; Length 3240;
Best Local Similarity 54.0%; Pred. No. 2.5e-83;
Matches 1442; Conservative 0; Mismatches 1187; Indels 39; Gaps 7;
Qy 662 CGACGACTTACCAACCTGCTGGAGCAGGCGGAGAGAGGTTGAAAAAGCTCCTTGAACA 721
Db 61 CGCCGCCCGCCCGCCGCCGCCGCGCGAGCTCAGATTAAAGAAACTAATTGATCA 120
Qy 722 AGAAAAAGCTTACCAAGCCCCGCAAGAAAAAGGAAACGCTAAGCGGCTCAACAACTCG 781
Db 121 AGAACTGCGTCCCAGGAAAAGAGGAGCAAGAAAAGGAGAGAGGATCAAGACCCCTGAA 180
Qy 782 AGATGAGCTTGTGAAGCTCAAGTCCCTTCGCCCTCATGTTGGTGGACGAGGCGAGATGCA 841
Db 181 GGAGGAGCTGACCAAGCTGAAGTCCCTTTGCTTTGATGGTGGTGAACAACAACCGCT 240
Qy 842 CATCGACCAACTGGGCTGCAGAGTCAGAAAAGTCCAGGACCTCACTCAGAAAGCTGAGGGA 901
Db 241 TACAGCTCAACTCGCCCTTCAAAGACAGAAAAATCCAAAGCTCTGACCACAAGTGCAAGGA 300

Qy 902 GGAGGAAGAAAAAAGCTCAAAGCGGTCACTTACAAAATCCAAGGAAGACCGCAGAAAGCTGCT 961
Db 301 AACACAGGGTAAACTAGCCCTTGTGTAGGCCAGAGCTCAGGAAGAAGAGCAGAAAGGCAAC 360
Qy 962 CAAGTTAGAAGTGGACTTCGAACACAAAGCCTCGAGGTTTTCAGAGGACGACGAAGAGAT 1021
Db 361 CAGACTAGAAAAAGAACTGCAGACACAGACCACAGAGTTTCACCAAAACCAAGACAAAAT 420
Qy 1022 GAACGCCAAATTTGGCGAATCAAGAATCTCACAAACCGGCAACTTCGACTCAAACCTGTTGG 1081
Db 421 TATGGCGAAGCTCACCAATGAGGACAGTCAAAAATCGCCAGCTCCGACAGAAAGCTGGCAGC 480
Qy 1082 CTTATCGCAAAAGGATTGAGGAGCTGGAAGAGAGACCAATFAAAAAGCCTTCAGAAAGGCAGAGGA 1141
Db 481 ACTCAGTCGGCAAAATTGATGAGTAAGAAAGAGACCACAGGTCTTTTAAGAAAAAGCCGAGGA 540
Qy 1142 AGAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAACCTCCAGTCTCAT 1201
Db 541 AGAGCTGCAGGATATAAAAAGACAAAATAAACCAAGGGGAGAATATGGAAGTCTTGGCATCAT 600
Qy 1202 GCGGAAGTGGAGAGTCTGCGCAAGCGCGTGTCTTGAGATGGAGGGCAAGGATGAAGAGAT 1261
Db 601 GGATGAAGTGGATGAGCTCAGGAAGCGCGTGTGACATGGAAGGGAAGAGATGAAGAGCT 660
Qy 1262 CACGAAGACCGAGGCCAGTCCCGGAGCTGGAAGAAAGCTCCAAGAGGAAGAACACCA 1321
Db 661 CATAAAAATGGAGGACGAGTGCAGAGACCTCAATAAGAGGCTAGAGAAAGAGACAGTACA 720
Qy 1322 CAGCAAGGAACTTAGACTAGAAGTGGAGAAAGCTGCAGAAAGGATGTCTGAGCTGGAGAA 1381
Db 721 GAGTAAAGACTTCAAACTAGAGGTTGACAAAACCTTAGTGTAAAGGATCACAGCTCTGGAAAA 780
Qy 1382 GCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGA 1441
Db 781 ATTAGAAGATGCTTTTAGACAAAAAGCAAAACAAAGATGCTACTCTCTGAAATGCAATTTAGA 840
Qy 1442 GAAGGAGAAGAACCTTAACCAAAGACCTGCTGAACAGAGCTGGAGGTGGTCAAGAGTCAGT 1501
Db 841 AAAAGAAAAGATGACCAACAAAGCAATTGTCTGAAGAGCTGGAGAGTTTAAATGCCAGAA 900
Qy 1502 TAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAAGCCGAGTTTAAGCCTCAAAGATGA 1561
Db 901 CAAAGAGCTAGAAGCCATTGAAAGTCGGCTAGAAAAAGACAGAAATCACACATAAAGGATGA 960
Qy 1562 CCTTACAAAGCTGAAGTCTTCACTGTGTATGCTGGTGGATGAGAGAAAAAATATGATGGA 1621
Db 961 TTTTAACCTAACTGAAAACTTTAACTGTGTATGCTTTGTAGATGAACGCAAAAACAATGAGTGA 1020
Qy 1622 GAAATAAAGCAAGAGAGAGGAAGTGGATGGGTGTAATAAAAACCTTTAAGGTGGAGCA 1681
Db 1021 AAAATTAAAGCAACAGAAAGATAAGTTACAAAGCACCACTCTCAGCTCCAAGCAGAGCA 1080
Qy 1682 GGGAAAAAGTCATGGATGTGACGGAAAAAGCTAATCGAGGAAAAGCAAGAAAGCTTTTAAAACT 1741
Db 1081 AAATAAAGTAAACACAGTGACTGAGAAAGTTAAATTGAGGAAAACCAAGAGGGCACTCAAGTC 1140
Qy 1742 CAAATCTGAAATGGAGGAAAAGGAGTACAGTCTGACAAAAGGAGAGGGATGAGCTGATGGG 1801
Db 1141 CAAAACGTGATGCAGAAAGAAAGATGTACAGTGTAAACCAAGGAGAGAGATGACCTCAGAAA 1200
Qy 1802 TAAACTGAGGAGCGAAGAAAGAGGTCTCTGTGAAGTGAAGTGCAGTGCAGTGTAGACTTACTAAA 1861
Db 1201 TAAACTGAAAGCAGAAAGAAAGAGAAAGGACATGATCTCTTGTCAAAAAGTTACTATATTGAA 1260
Qy 1862 GAACGGGCTTGA-----TGGCATAGAGGAGGTTAGAAAAGGGGAAATAAAACCG 1906
Db 1261 AAACAGGCTTCAGTCACCTGGAGCAATTTGAGAAAGATTTTGTAAAAAACAATTTAAACCA 1320
Qy 1907 AGGTAGGTCGTGCAAGGGGTCTGAGTTACCTGCCCGGAAGACAAATAAGATCAGAGAACT 1966
Db 1321 AG---ACTCCAGTAAGTCCACAGCAGCATTTACACCAAGAAAAACAATAAGATTAAAGAGCT 1377

QY 1967 AACGCTTGAAATCGAGAGACTGAAGAAAACGGCTTCAGCAGTGTGGAGGTGGAGGGGGA 2026
Db 1378 CTCTCAAGAAGTGGAAAATCTGAAACTGAAGCTAAAGGATATGAAAGCCATTCAGGATGA 1437
QY 2027 CTTGATGAAGACCGAGGACGAATATGACCAAGTTGGAGCAGAAGTTTCAGAACCGAGCAGGA 2086
Db 1438 CCTCATGAAAACAGAGGATGAGTACGAGACCTTAGAACGGAGGTATGCTAACGAGAGAGA 1497
QY 2087 TAAGGCNAACCTTCCTCTCCAGCAGCTCGAGGAATCAAAACACCAAAATGGCCAAGCACAA 2146
Db 1498 CAAGGCTCAGTTTCTGTCTCAGGAGCTGGAACATGCAAGATGGAGCTTGCACCAAGTACAA 1557
QY 2147 AGCCATAGAGAAAAGGGAGGCGCGTGAGCCAGGAAGCCGAACCTGCGACACAGGTTTCGGCT 2206
Db 1558 GTTGGCCGAGAAGCAGAGTCC--AGCCACGAGCAGTGGCTCTTTAGGAGGCTCCAGGA 1614
QY 2207 GGAGGAGGCTAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAGGAGAAGATCCA 2266
Db 1615 AGAAGAAGCAAAATCAGGGCACCTGTCCAGAGAAAGTGATGCATTAAGAGAGAAAATTCA 1674
QY 2267 CGAGCTGATGAACAAAGGAAGACACAGCTGTCTCAGTCCAAAGTCGACTATTCGGTCTCTCA 2326
Db 1675 TGAGTACATGGCCACAGAGGACCTAATATGTACCTCCAAGGAGACCACCTCCCTTCGTGA 1734
QY 2327 GCAAAGATTTATGGAAGAAGAAACTAAGAAACAAGAACATGGGAGGGAGGTCTCAATCT 2386
Db 1735 AAAGAACTAAACCAACAAGAAAACAGGAACAGAGATTTGGGAAGAGAGATTGAAAAACCT 1794
QY 2387 GACCAAGGAGCTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGG 2446
Db 1795 CACTAAAGAGTTAGAAAGGTATAGGCACCTTTAGTAAGAGCCTCCGGCCTAGTCTCAACGA 1854
QY 2447 CCGAAGGATGGTGGACGTGCCTGTGGCCTCCACTGGGGTGCAGACCGAGGCGGTGTGGG 2506
Db 1855 AAGAAGAATCTCTGACCCGCAAGTATTTCTAAGAAGTTCAAACGGAAAGCAGCAGACAG 1914
QY 2507 GGATGCTGCGGAGGAGGAGACCCCGCTGTGTTCATTCGCAAAATCCTTCCAGGAGGAAA 2566
Db 1915 TGAGCCACAGACTACAAGAGCCTCATTCCTCTGGAAGAGCAGTGATCAATGGTCAGTT 1974
QY 2567 TCACATCATGAGTAATCTTCGACAGGTAGGCTGAAGAAACCCATGGAACGGTCTCGGT 2626
Db 1975 CTATGAGGAGAATGAGGACCAGGATGACGACCCTAATGA-----GGAGGAGTCTGTGCT 2028
QY 2627 CCTCGACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTGGATTCTTG 2686
Db 2029 GTCCTTCAGATGCAGCCAGTCTTCTCTCTTCCTATGAACAGGAACACTATGATCCCTG 2088
QY 2687 GATGAGAAAAGAGAAAACGGTCTCTCCACTCCGAGGAGAAAGGGCCCCAGGCCAAACCA 2746
Db 2089 GATGAAATCCAAGAGGGCCATCTCTCAGAAAT---GGAAAAATACAACTAAGTCCAATGG 2145
QY 2747 GGGTGCAGGGCACCCCGGGGAGCTGGTCTCTAGCAACAAAGCAGGGCCAGCCCTACACAT 2806
Db 2146 CAACTTTGTACAACCTGGAGATCTTGTCTTAAGTCACACACACCTGGGCAGCCACTTCACAT 2205
QY 2807 CCGTGTGACACCCAGATCATGAGAACAGCACCTGCCACCCTGGAGATCAACAGCCCCACATC 2866
Db 2206 AAAGGTTACTCCAGACCACATTCAAAATACAGCTACTCTTGAAATCACAAGTCCAACCAC 2265
QY 2867 TGAAG-----AGTTTTTCTTAGTACCACCGTCACTTCTACCTTAGGCAACCAAGAAC 2920
Db 2266 AGAGAGTCTCACTCTACACCAGCACAGCGGTGATACCAAACTGTGGCACCCCAAGCA 2325
QY 2921 AAGAATAACCATTTATCCCATCACCCAATGTCTATGTCGCAAAAGCCCAAG---TGCAGA 2977
Db 2326 AAGGATTACCATTTCTCCAAAATGCCTCTATAACACCAATTAAGTCCAAAAGCTCTACAGA 2385
QY 2978 TCCTACTCTCGGCCAGAACGAGCCATGTCCCCTGTCAAGATTACTACTATTTCCAGAGA 3037
Db 2386 AAGCCTTATGAACCTTAGAGCAAAAGCATGTCCCAGTTACCATGGCAACCTTTGGCAGAGC 2445
QY 3038 GAAGAGCCCGAAGGTGGAAGGAGCGCCCTTTGCGGACAGGCGCTGTCATCCCCCATCCAAAT 3097

Db 2446 ACAGACCCCGAGAGTCTGTGGTTCTGTGACTCCAGAACGGACAATGTACCTATTTCAGGT 2505
QY 3098 CATGACGGTGTCAACATCTGCAGCTCCCACTCCCAATCGCTGTCTCTCCTGAATCTCAGGA 3157
Db 2506 TTTGGCTATGACTGGTTTCACTAGTTCTCTCTGAGCAAGGCTGCTCCCCAGAGCCGATAGA 2565
QY 3158 AGTGCCTATGGGAAGGACTATCCTCAAAGTCAACCCCGGAAAAACAACCTGTTCCAGCCCC 3217
Db 2566 AATCAGCGCCCAAGCATGCAATTTTCAGAGTCTCCCAGACCGGCAGTCATCATGGCAGTT 2625
QY 3218 CGTGGGAAGTACAACTCCAATGCTAATATCATCATCACCGGAAGACAATAAATTCACAT 3277
Db 2626 TCAACGTTCAAACAGTAACAGCTCAAGTGTGATAACTACTGAGGATAATAAATCCACAT 2685
QY 3278 TCACCTGGGTTCTCAGTTTAAGCGATCT 3305
Db 2686 TCACCTAGGAAGTCTCTACATGCAAGCT 2713

RESULT 6
AI176515/c
LOCUS
DEFINITION
AI176515
EST220101 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
ROVBS89 3' end, mRNA sequence.
ACCESSION
AI176515
VERSION
AI176515.1 GI:3727153
KEYWORDS
EST.
SOURCE
Rattus sp.
ORGANISM
Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 618)
Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
Gene Index
Unpublished (1998)
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1. .618
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="ATCC (inhost):2031381"
/db_xref="taxon:10118"
/clone="ROVBS89"
/clone_lib="Normalized rat ovary, Bento Soares"
/note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI;
Site_2: NotI"

ORIGIN

Query Match 14.1%; Score 615.4; DB 9; Length 618;
Best Local Similarity 99.8%; Pred. No. 2.9e-82;
Matches 616; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3698 AGGGGGAAGGGCTGAGGGCAGTGGCTAAGGGGGTATGTTGTAAGGATGCTACTGTGCA 3757
Db 617 AGGGGGAAGGGCTAAGGGCAGTGGCTAAGGGGGTATGTTGTAAGGATGCTACTGTGCA 558
QY 3758 GTGGAACAAACCTTCTCTGTGCCAACCCCTTCTTGACTACTAATAATTAAGTTTAAA 3817
Db 557 GTGGAACAAACCTTCTCTGTGCCAACCCCTTCTTGACTACTAATAATTAAGTTTAAA 498
QY 3818 TATCTTGTATTATAAAATAAACCATTTAATAGCATGCAACCCCTCCCATTTTGTGCATCT 3877
Db 497 TATCTTGTATTATAAAATAAACCATTTAATAGCATGCAACCCCTCCCATTTTGTGCATCT 438

```
QY 3878 GTTTCATGCAGGGGAATAGAAATTAATTAGCAGAAATTTCTGTTTGGTGAATGTTCTGTGTG 3937
Db |||||||
QY 437 GTTTCATGCAGGGGAATAGAAATTAATTAGCAGAAATTTCTGTTTGCIGAAATGTTCTGTGTG 378
Db |||||||
QY 3938 AAGATGTTGGTCCAGTTCAGTCTTACTTCTAGCATGTGGCCCCCATTCAAAGGTAGCTCACG 3997
Db |||||||
QY 377 AAGATGTTGGTCCAGTTCAGTCTTACTTCTAGCATGTGGCCCCCATTCAAAGGTAGCTCACG 318
QY 3998 AGTTGTGAAGCCCTCAATATCGTCACCGAGAGATTTGAGGACCAATACATATGCTCC 4057
Db |||||||
QY 317 AGTTGTGAAGCCCTCAATATCGTCACCGAGAGATTTGAGGACCAATACATATGCTCC 258
QY 4058 CAAAGGTGGCTCCCAATTTTCTCTAAATTTGTAAGCCCACTTAAATAGACTCAGTTCTGTGA 4117
Db |||||||
QY 257 CAAAGGTGGCTCCCAATTTTCTCTAAATTTGTAAGCCCACTTAAATAGACTCAGTTCTGTGA 198
QY 4118 TTTTCTTTTCCAAAAAATAATTTTGAAATAGGACAGAGTTTAAACAGTTGTCAATTTG 4177
Db |||||||
QY 197 TTTTCTTTTCCAAAAAATAATTTTGAAATAGGACAGAGTTTAAACAGTTGTCAATTTG 138
QY 4178 CACTATCAAGCCCATGAGTTTGATATATGGGTTATAAGAAAGAATACTTTCAGAGCTATC 4237
Db |||||||
QY 137 CACTATCAAGCCCATGAGTTTGATATATGGGTTATAAGAAAGAATACTTTCAGAGCTATC 78
QY 4238 ACAGGGTCTCTAAACTTTTGAAAAACAAAAAGCCCCCTAATATGACCTCAGGAAACAATTT 4297
Db |||||||
QY 77 ACAGGGTCTCTAAACTTTTGAAAAACAAAAAGCCCCCTAATATGACCTCAGGAAACAATTT 18
QY 4298 GAACATGAAATAAAATG 4314
Db |||||||
QY 17 GAACATGAAATAAAATG 1
```

```
RESULT 7
BC030632
LOCUS BC030632.1 GI:21040469
DEFINITION Mus musculus, Similar to KIRAI275 protein, clone IMAGE:1245919, mRNA.
ACCESSION BC030632
VERSION BC030632.1
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
```

```
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 777)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
```

```
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Marcello Bento Soares, Ph.D.
cDNA Library Preparation: Soares Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbiology.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
```

```
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 66 Row: 9 Column: 21
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis
This clone has the following problem: frame shifted.
Location/Qualifiers
```

FEATURES

```
source 1. .777
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1245919"
/tissue_type="Mammary gland"
/clone_lib="Soares_mammary_gland_NbMMG"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac"

ORIGIN
```

```
Query Match 14.1%; Score 615.4; DB 11; Length 777;
Best Local Similarity 93.0%; Pred. No. 2.7e-82;
Matches 656; Conservative 0; Mismatches 46; Indels 3; Gaps 1;

QY 56 GAGTCGACAAACAGGTGGGAATGAGATCACGAAATCAAGGTGGAGAAAAGTTCATCTAACGG 115
Db |||
QY 60 GAGATGAGTAGCGGTGGGAATGAGATCTCGAAACCAAGGTGGAGAAAAGTTCATCTAATGG 119
QY 116 GCATGCTCTCCTGCCCAAGTCTCCATCATCATCAGCAGTGGTGGTAAGGCCCCCTCAGA 175
Db |||
QY 120 GCATGCTCTCCTGCCCAAGCCTCCATCATCATCAGCAGTGGTGGTAAGGACCCTCAGA 179
QY 176 AGATGCAAAAAAGAACAAAGGCCAATCGGAA--GGAGAGGATGTCTATGGCTTCCGGAAC 232
Db |||
QY 180 AGATGCAAAAAAGAACAAAGGCCAATCGGAAAGGGGAGGATGATGTCTATGGCTTCCGGAAC 239
QY 233 TATCAAAAGGCACCTCAAAACCATCTCGAGAAAAGTGAGAAAAAGACTAAGAAGTCTGTGGA 292
Db |||
QY 240 CGTCAAAAGCCACCTCAAGCCATCAGGAGAGAGTGAGAAAAAGACTAAGAAGCCTCTGGA 299
QY 293 GTTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGAGTTGCAGGCTCG 352
Db |||
QY 300 GTTATCCAAGGAGGACCTCATCCAGCTCCTGAGCATCATGGAAGGGAGCTGCAGGCTCG 359
QY 353 AGAAGATGTATCCACATGCTGAGGACAGAGAGAAAAACCAAGCCCCGAGGTTCTGGAGGCACA 412
Db |||
QY 360 AGAAGATGTATCCACATGCTGAAGACGGAGAGAACCACCAAGCCCCGAGGTTCTGGAGGCGCA 419
QY 413 CTATGGATCTGCAGAAACCTGAGAAAAGTGCTTCGGGTCTCTGCACCGAGATGCCATCCTTGC 472
Db |||
QY 420 TTATGGATCTGCAGAAACAGAGAAAAGTGCTTCGAGTCTCTGCACCGGATGCCATCCTTGC 479
QY 473 TCAAGAGAAAGTCCATAGGAGAAGACGTCTATGAGAAACCTATCTCAGAGTGGACAGACT 532
Db |||
QY 480 TCAGGAGAAGTCCATAGGAGAAGATGTCTATGAGAAACCCATCTCAGAGCTGGACAGACT 539
QY 533 GGAGGAAAAGCAGAACGAGACGCTACCCCGCATGCTTAGAGCAGCTGCTCTGGCTGAGAA 592
Db |||
QY 540 GGAGGAAAAGCAGAACGAGACATACCCCGCATGCTGGAGCAGCTGCTGTTGGCTGAGAA 599
QY 593 GTGTCAACAGGCGCACCGTGTACGAGCTGGAGAACGAGAAAGCACAAGCACACTGACTACAT 652
Db |||
QY 600 GTGTCAACAGGCGCACCGTGTACGAGCTGGAAAAACGAGAAAGCACAAGCACCCGACTACAT 659
QY 653 GAACAAGAGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGAGAGAGGTTGAAAAAGCT 712
Db |||
QY 660 GAACAAGAGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGAGAGAGGTTGAAAAAGCT 719
QY 713 CCTTGAACAAGAAAAAGCTTTACCAAGCCCGCAAGAAAAAGGAAAA 757
Db |||
QY 720 CCTTGAGCAAGAAAAAGGCTTACCAAGCCCGCAAAAAAAGAAAAA 764
```

```
RESULT 8
CB581318
LOCUS CB581318
DEFINITION AMGNNUC:NRDGI-00062-F8-A nrdgl (10855) Rattus norvegicus cDNA clone nrdgl-00062-f8 5', mRNA sequence.
ACCESSION CB581318
VERSION CB581318.1
KEYWORDS EST.
GI:29525359
```

SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 606)
AUTHORS Angen EST Program.
TITLE Angen Rat EST Program
JOURNAL Unpublished (2003)
COMMENT Contact: Dan Fitzpatrick
Angen, Inc
One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00062 row: f column: 8.
FEATURES Location/Qualifiers
source 1..606
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="nr01-00062-f8"
/tissue_type="Dorsal Root Ganglia"
/clone_lib="nr01 (10855)"
/note="vector: pSPORT1; Site_1: Sall; Site_2: NotI; rat
dorsal root ganglia"
ORIGIN
Query Match 13.9%; Score 606; DB 14; Length 606;
Best Local Similarity 100.0%; Pred. No. 7.3e-81;
Matches 606; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2018 GGAGGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAGCAGAAAGTTCAGAAC 2077
Db 1 GGAGGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAGCAGAAAGTTCAGAAC 60
QY 2078 CGAGCAGGATAAGGCCTTCTCTCCAGCAGCTCGAGGAAATCAAACACCAATGGC 2137
Db 61 CGAGCAGGATAAGGCCTTCTCTCCAGCAGCTCGAGGAAATCAAACACCAATGGC 120
QY 2138 CAAGCACAAAGCCATAGAGAAAGGGAGGCCGTGAGCCAGGAAGCCGAACCTGGCACACAG 2197
Db 121 CAAGCACAAAGCCATAGAGAAAGGGAGGCCGTGAGCCAGGAAGCCGAACCTGGCACACAG 180
QY 2198 GTTTCGGCTGGAGGCTAAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAGGA 2257
Db 181 GTTTCGGCTGGAGGCTAAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAGGA 240
QY 2258 GAAGATCCACGAGCTGATGAACAAGGAGACCAGCTGCTCAGCTCCAAGTCGACTATTC 2317
Db 241 GAAGATCCACGAGCTGATGAACAAGGAGACCAGCTGCTCAGCTCCAAGTCGACTATTC 300
QY 2318 GGTCTTTCAGCAAGATTATGGAAGAAGAACTAAGAAACAAGAACATGGGGAGGGAGGT 2377
Db 301 GGTCTTTCAGCAAGATTATGGAAGAAGAACTAAGAAACAAGAACATGGGGAGGGAGGT 360
QY 2378 CCTCAATCTGACCAAGGAGCTAGAGCTTTCGAAGCGCTACAGCCGAGCTCTCAGGCCGAG 2437
Db 361 CCTCAATCTGACCAAGGAGCTAGAGCTTTCGAAGCGCTACAGCCGAGCTCTCAGGCCGAG 420
QY 2438 TGGGAACGGCCGAGAGTGGTGGACGTGCCCTGTGGCTCCACTGGGGTGCAGACCGGAGGC 2497
Db 421 TGGGAACGGCCGAGAGTGGTGGACGTGCCCTGTGGCTCCACTGGGGTGCAGACCGGAGGC 480
QY 2498 GGTGTGCGGGGATGTGCGGAGGAGGAGACCCCGCTGTGTTTCATTCGCAATCCTTCCA 2557
Db 481 GGTGTGCGGGGATGTGCGGAGGAGGAGACCCCGCTGTGTTTCATTCGCAATCCTTCCA 540
QY 2558 GGAGGAAAATCAGATGAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCATGGAACG 2617
Db 541 GGAGGAAAATCAGATGAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCATGGAACG 600
QY 2618 GTCCTC 2623
Db 601 GTCCTC 606

RESULT 9
AUI30457
LOCUS AUI30457 NT2RP3 Homo sapiens cDNA clone NT2RP3000868 5', mRNA linear EST 01-AUG-2002
DEFINITION AUI30457 NT2RP3 Homo sapiens cDNA clone NT2RP3000868 5', mRNA
sequence.
ACCESSION AUI30457
VERSION AUI30457.1 GI:10990811
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 844)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES Location/Qualifiers
source 1..844
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP3000868"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/clone_lib="NT2RP3"
/note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"
ORIGIN
Query Match 13.3%; Score 581.6; DB 9; Length 844;
Best Local Similarity 81.1%; Pred. No. 2.9e-77;
Matches 685; Conservative 0; Mismatches 159; Indels 1; Gaps 1;
QY 1719 GAAAGCAAGAAGCTTTTAAAACTCAAATCTGAAATGGAGGAAAGGAGTACAGTGTGACA 1778
Db 1 GNAAGTAAGAACTTTTAAAACTAAATCTGAAATGGAGGAAAGATATACAACTGTGACA 60
QY 1779 AAGGAGAGGGATGAGCTGATGGTAAACTGAGGAGCGAAGAAAGGTCTGTGAACTG 1838
Db 61 AGAGAAAGAGATGAGTTGATAGGCCAAATTGAAAGAGTGAAGAAAGAAATCCTCTGAATTA 120
QY 1839 AGCTGCAGTGTAGACTTACTAAAGAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGGGAA 1898
Db 121 AGCTGCAGTGTGACTTACTAAAGAAGAGACTTGATGGTATAGAGGAAGTGGAAAGAGAA 180
QY 1899 ATAAACCGAGGTAGTCGTGCAAGGGGTCTGAGTTCACCTGCCCGAAGACATAAGATC 1958
Db 181 ATAACAAGAGGAAGGTCACGAAAGGGTCTGAGTTCACCTGCCCGAAGATATAAGATT 240
QY 1959 AGAGAACTAACGCTTGAAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTGGTG 2018
Db 241 AAGGAACTAACACTTGAAATTGAGAGACTGAAGAAACGTCTCCAACAATTGGAAGTGGTC 300
QY 2019 GAGGGGACTTGATGAAGACCGAGGACGAAATATGACCAGTTGGAGCAGAGATTTCAGAAC 2078
Db 301 GAAGGGGATTGTGATGAAGACAGAGATGAGTATGATCAGCTGGAACAGAAATTTAGAACT 360
QY 2079 GAGCAGGATAAGGCAAACTTCTCTCCAGCAGCTCGAGGAAATCAAACACCAATGGCC 2138

Db 361 GAGCAGGATAAGCGTAACCTTCCTCTCTCAACAACTAGAGGAGATCAAGCACCAAATTGCC 420
QY 2139 AAGCACAAAGCCATAGAGAAAGGGGAGGCCGTGAGCCAGGAAGCCGAACTGGCACACAGG 2198
Db 421 AAGAAATAAGCAATAGAGAAGGCTGAGGTTGTGAGCCAGGAAGCTGAAGTGAAGACACAGA 480
QY 2199 TTTCGGCTGGAGAGGCTAAAGTCGTGATTATACAGGCCGAGGTGAGGCTCTCAAGGAG 2258
Db 481 TTTCGGTTGGAAGAAGCTAAAGTCGAGACTTAAAGCCGAAGTACAAGCTCTTAAAGAG 540
QY 2259 AAGATCCACGAGCTGATGAACAAAGGAAGACAGCTGTCTCAGCTCCAAGTCGACTATTTCG 2318
Db 541 AAGATTCACGAATTAATGAACAAAGAAGATCAGCTTCTCAGCTCCAGGTAGATTATTCT 600
QY 2319 GTCCTTCAGCAAAGATTATGGAAGAAGAAACTAAAGAACRAAGAACATGGGGAGGGAGGTC 2378
Db 601 GTACTTCAACAAAGATTATGGAAGAAGAAATAAGAACAAAAACATGGGCGANGAGGTT 660
QY 2379 CTCAATCTGACCAAGGAGCTAGAGCTTTCCAAAGCGCTACAGCCGAGCTCTCAGGCCGAGT 2438
Db 661 CTCAATCTGACCAAGAGATTGGAGCTTTCCAAAGCGCTACAGCAGAGCTCTTA-GCCCCAGT 719
QY 2439 GGGAAACGGCCGAGGATGGTGGACGTGCTGTGGCCTCCACTGGGGTGCAGACCGAGGCG 2498
Db 720 GTGAATGGAAGAAGAATGGTGGATGTTCTCTGTGACGTCACTGAGTCCAAACTGATGCA 779
QY 2499 GTGTGCGGGGATGCTGCGGAGGAGGAGACCCCGGCTGTGTTCAATTCGCAAAATCCTTCCAG 2558
Db 780 GTCCAGCGGTGAANCACAGANGAAGAAACGCCAGCTGTATTATACCGAAATCCTTNCAG 839
QY 2559 GAGGA 2563
Db 840 GAAGA 844

RESULT 10
BF562820
LOCUS
DEFINITION
UI-R-BJ0p-aio-f-08-0-UI.r1 UI-R-BJ0p Rattus norvegicus cDNA clone
UI-R-BJ0p-aio-f-08-0-UI 5', mRNA sequence.
BF562820
VERSION
BF562820.1 GI:11672550
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 583)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
AUTHORS
Normalization and subtraction: two approaches to facilitate gene
TITLE
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477
PUBMED
8889548
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LLNL (info@image.llnl.gov). IMAGE ID= 1797667
Seq primer: M13 Forward.
Location/Qualifiers
1..583
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"

FEATURES
source

/db_xref="taxon:10116"
/clone="UI-R-BJ0p-aio-f-08-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-BJ0p"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ0p
library is a subtracted library derived from the UI-R-AAL,
UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and
UI-R-AG1 libraries. These libraries represent tissues from
rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
canal at 15 dpc, and ventricle at 13 dpc. The tag is a
string of 5-6 nucleotides present between the Not I site
and the oligo-dT track. The library was constructed as
described by Bonaldo, Lennon and Soares, Genome Research
6: 791-806, 1996."

ORIGIN

Query Match 13.2%; Score 576.2; DB 10; Length 583;
Best Local Similarity 99.5%; Pred. No. 2.1e-76;
Matches 578; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 3747 CTACTGCTGCAGTGGAAACAAACCTTCTCTGTGCCAACCCCTTTCCTTGTTACTACTAATT 3806
Db 3 CGAGGCTGCAGTGGAAACAAACCTTCTCTGTGCCAACCCCTTTCCTTGTTACTACTAATT 62
QY 3807 TAAGTTTTAAATATCTTGTTTTATAAAATAACCATTTAATAGCCATGCACCCCTCCCAT 3866
Db 63 TAAGTTTTAAATATCTTGTTTTATAAAATAACCATTTAATAGCCATGCACCCCTCCCAT 122
QY 3867 TTTGTGCATCTGTTTCAATGCAGGGGAATAGAAATTAATTAGCAGAATTTCTGTTGCTGA 3926
Db 123 TTTGTGCATCTGTTTCAATGCAGGGGAATAGAAATTAATTAGCAGAATTTCTGTTGCTGA 182
QY 3927 ATGTTCTGTTGAAGATGTTGGTCAGTTCAGTTTCAGTTTACTTCTAGCATGTGGCCCCATTCAA 3986
Db 183 ATGTTCTGTTGAAGATGTTGGTCAGTTCAGTTTCAGTTTACTTCTAGCATGTGGCCCCATTCAA 242
QY 3987 GGTAGCTCACGAGTGTGAAGCCCTCAATATCGTCACCGGAGAGATTTGAGGACCACATT 4046
Db 243 GGTAGCTCACGAGTGTGAAGCCCTCAATATCGTCACCGGAGAGATTTGAGGACCACATT 302
QY 4047 ACATATGCTCCCAAAGCTGGTCCCAATTTTCTAATTTGTAAGCCAACTTTTATFAGACT 4106
Db 303 ACATATGCTCCCAAAGCTGGTCCCAATTTTCTAATTTGTAAGCCAACTTTTATFAGACT 362
QY 4107 CAGTTCTGTGATTTTTTTTCCAAAAAATAATTTTGAATAGGACAGAGTTTAAACAG 4166
Db 363 CAGTTCTGTGATTTTTTTTCCAAAAAATAATTTTGAATAGGACAGAGTTTAAACAG 422
QY 4167 TTGTCATTTTGCACTATCAAGCCATGAGTTTGATATATGGTTTATAAGAAAAAGATACTT 4226
Db 423 TTGTCATTTTGCACTATCAAGCCATGAGTTTGATATATGGTTTATAAGAAAAAGATACTT 482
QY 4227 TCAGAGCTATCACAGGGTCTCTAAACTTTTGAAAAACAAAAGCCCTTAATATGACCTCA 4286
Db 483 TCAGAGCTATCACAGGGTCTCTAAACTTTTGAAAAACAAAAGCCCTTAATATGACCTCA 542
QY 4287 GGAAAACAATTTGAACATGAATAAAATGGAATGAACTGTG 4327
Db 543 GGAAAACAATTTGAACATGAATAAAATGGAATGAACTGTG 583

RESULT 11
BE095605/c
LOCUS
DEFINITION
UI-R-B01-apo-a-05-0-UI.s1 UI-R-B01 Rattus norvegicus cDNA clone
UI-R-B01-apo-a-05-0-UI 3', mRNA sequence.
BE095605
ACCESSION
BE095605.1 GI:8486536
VERSION
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)

BE095605 582 bp mRNA linear EST 12-JUN-2000
UI-R-B01-apo-a-05-0-UI.s1 UI-R-B01 Rattus norvegicus cDNA clone
UI-R-B01-apo-a-05-0-UI 3', mRNA sequence.

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 582)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel.: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized corpus-striatum library cDNA Library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.
FEATURES source
LOCUS 1. .582
DEFINITION /organism="Rattus norvegicus"
ACCESSION /mol_type="mRNA"
KEYWORDS /strain="Sprague-Dawley"
SOURCE /db_xref="taxon:10116"
ORGANISM /clone="UI-R-B01-apo-a-05-0-UI"
REFERENCE /dev_stage="adult"
AUTHORS /lab_host="DH10B (Life Technologies)"
TITLE /clone_lib="UI-R-B01"
JOURNAL note="Vector: pT7T3D-Pac (Pharmacia) with a modified
MEDLINE polylinker; Site 1: Not I; Site 2: Eco RI; The library
PUBMED (UI-R-B01) is a subtracted library derived from a mixture
COMMENT of the following tissues: thalamus, cerebellum,
hypothalamus, medulla, pons, midbrain, cerebral cortex,
corpus striatum and hippocampus. For a detailed
description of the library from which this clone was
derived, please visit our web site at
ratest.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG TISSUE=corpus-striatum
TAG_LIB=UI-R-B01
TAG_SEQ=CTAGG"
ORIGIN
Query Match 13.1%; Score 573.2; DB 10; Length 582;
Best Local Similarity 99.3%; Pred. No. 5.9e-76;
Matches 575; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 179 TGCAAAAAGAACAGGCGCAATCGGAAGGAGGAGGATGTCTATGGCTTCGGAACTATCAA 238
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
582 TGCAAAAAGAACAGNCCAAATCGGAAGGAGGAGGATGTCTATGGCTTCGGAACTATCAA 523
QY 239 AAGGCACCTCAAAACCATCTGGAGAAAGTGAGAAAAGACTAAGAAGTCTGTGGAGTTATC 298
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
522 AAGGCACCTCAAAACCATCTGGAGAAAGTGAGAAAAGACTAAGAAGTCTGTGGAGTTATC 463
QY 299 CAAGGAGGACCTCATCCAGCTCCTGAGTATCATGTGAAGGGGAGTTGCAGGCTCGAGAAGA 358
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
462 CAAGGAGGACCTCATCCAGCTCCTGAGTATCATGTGAAGGGGAGTTGCAGGCTCGAGAAGA 403
QY 359 TGTCAATCCACATGCTGAGGACAGAGAAAAACCAAGCCCGAGGTTCTGGAGGCACACTATGG 418
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
402 TGTCAATCCACATGCTGAGGACAGAGAAAAACCAAGCCCGAGGTTCTGGAGGCACACTATGG 343

QY 419 ATCTGCAGAACCTGAGAAAGTGTCTCGGGTCTCTGCACCGAGATGCCATCCTTGCTCAAGA 478
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
342 ATCTGCAGAACCTGAGAAAGTGTCTCGGGTCTCTGCACCGAGATGCCATCCTTGCTCAAGA 283
QY 479 GAAGTCCATAGGAGAGACGCTCTATGAGAAAACCTATCTCAGAGCTGGACAGACTGGAGGA 538
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
282 GAAGTCCATAGGAGAGACGCTCTATGAGAAAACCTATCTCAGAGCTGGACAGACTGGAGGA 223
QY 539 AAAGCAGAAGGAGACGTACCGCCGCATGCTAGAGCAGCTGCTGGCTGAGAAAGTGTC A 598
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
222 AAAGCAGAAGGAGACGTACCGCCGCATGCTAGAGCAGCTGCTGGCTGAGAAAGTGTC A 163
QY 599 CAGGCGCACCGTGTACGAGCTGGAGAACGAGAAAGCACAAGCACACTGACTACATGAACAA 658
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
162 CAGGCGCACCGTGTACGAGCTGGAGAACGAGAAAGCACAAGCACACTGACTACATGAACAA 103
QY 659 GAGCGACGACTTCACCAACCTGCTGGAGCAGGAGGAGAGAGGTTGAAAAAGCTCCTTGA 718
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
102 GAGCGACGACTTCACCAACCTGCTGGAGCAGGAGGAGAGAGGTTGAAAAAGCTCCTTGA 43
QY 719 ACAAGAAAAAGCTTACCAAGCCCCGCAAGAAAAAGGAAAA 757
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
42 ACAAGAAAAAGCTTACCAAGCCCCGCAAGAAAAAGGAAAA 4
RESULT 12
BG213104
LOCUS BG213104 785 bp mRNA linear EST 21-APR-2001
DEFINITION RST32708 Athersys RAGE Library Homo sapiens CDNA, mRNA sequence.
ACCESSION BG213104
VERSION BG213104.1 GI:13734791
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 785)
Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R.,
Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Boozer,S.,
Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K.,
Offenbacher,J., Danzig,J. and Ducar,M.
TITLE Creation of genome-wide protein expression libraries using random
activation of gene expression
JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE 21227151
PUBMED 11329013
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 342.
FEATURES Location/Qualifiers
source 1. .785
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_line="HT1080"
/clone_lib="Athersys RAGE Library"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
ORIGIN
Query Match 12.9%; Score 561.2; DB 12; Length 785;
Best Local Similarity 82.4%; Pred. No. 3.3e-74;
Matches 644; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY	522	CTGGACAGACTGGAGGAAAGCAGAACGAGACGTAACCGCCGCATGCTAGAGCAGCTGCTG	581
Db	1	CTGGACAGACTTGAGGAAACACAGAAAGAACCTACCGCGCATGCTAGAGCAGCTGTTG	60
QY	582	CTGGCTGAGAAAGTGTACACAGGCGCACCGTGTACGAGTGGAGAACGAGAAGCACACAGCAC	641
Db	61	CTGGCCGAGAAAGTGTATAGGCGCACCGTATACGAGTTAGAGAACGAGAAGCATAAACAC	120
QY	642	ACTGACTACATGAACAAGAGCGACGACTTCACCAACTGCTGGAGCAGGAGCGAGAGAGG	701
Db	121	ACTGACTACATGAACAAGAGCGACGACTTCACCAACTGCTGGAGCAGGAGCGGGAGAGG	180
QY	702	TTGAAAAAGCTCCTTGAAACAAGAAAAAGCTTACCAAGCCCGCAAAAGAAAAAGGAAACGCCT	761
Db	181	TTAAAAAAGCTCCTTGAAACAAGAAAAAGCGCTTATCAAGCCCGCAAAAGAAAAAGGAAATGCT	240
QY	762	AAGCGGTCAACAACTTCGAGATGAGCTTGTAAGCTCAAGTCAAGTCCITCGCCCTCATGTTG	821
Db	241	AAACGACTCAATAAACTAAGAGATGAGCTTGTTAAACTCAAACTCCTTTGCACCTCATGCTG	300
QY	822	GTGGACGAGGCGCAGATGCACATCGAGCAACTGGGCTTCAGAGTCAAGAAAGTCCAGGAC	881
Db	301	GTGGATGAAGAACAATGCACATTGAACAACCTTGGCTGCTCAAGCCAGAAAGTACAGGAT	360
QY	882	CTCACTCAGAAAGCTGAGGAGGAGGAGAAAGAAAACTCAAGCGGTCACTTACAAATCCAAG	941
Db	361	CTTACTCAGAAAGCTGAGGGAAGAGAGAGAGAGCTCAAGGCCATTACTTCCAAATCCAAA	420
QY	942	GAAGACGCCAGAAAGCTGCTCAAGTTAGAAAGTGGAATTCGAAACACAAGCCCTCGAGTTT	1001
Db	421	GAAGACAGACAGAAATTGCTCAAGTTAGAAAGTGGAATTTGAACACAAGGCTTCGAGGTT	480
QY	1002	TCCCAGGAGCACGAAGAGATGAACGCCAAATTGGCGAATCAAGAAATCTCACAAACCCGGCAA	1061
Db	481	TCTCAAGAGCATGAAGAGATGAACGCTAAACTGGCTAATCAAGAGTCTCACATTAGGCCA	540
QY	1062	CTTCGACTCAAACTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGAAAGACCAATAAA	1121
Db	541	CTTAGACTCAAGCTGGTGGCTTAACCCAAAGAAATCGAGGAGCTAGAAAAGACCCACAAA	600
QY	1122	AGCCTTCAAGAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAATTGCCAAAGGGAA	1181
Db	601	AATCTGCAAGAGGCAAGGAAGATCTTCCAGAATTAAAGAGATAAAAATTGCCAAAAGAA	660
QY	1182	TGTGGAATCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGGCTGTGATG	1241
Db	661	TGGGAAACTCTAGCCTAATGGGCAAGTGGAAAATTCTTCAAAGCCTGTGTTGGAATG	720
QY	1242	GAGGCAAGGATGAAGAGATCACGAAGACCGAGGCCAGTGCCTGGAGCTGAAGAAGAG	1301
Db	721	GAAGGTAAGATGAGGATATTCCCTTAACCGAACTCAAGGGATGGAATTTGAGGAGAAAC	780
QY	1302	CT 1303	
Db	781	CT 782	
RESULT 13			
BF398646/c			
LOCUS			
DEFINITION			
UI-R-BS2-ber-c-04-0-UI.s1 UI-R-BS2 Rattus norvegicus cDNA clone			
UI-R-BS2-ber-c-04-0-UI 3', mRNA sequence.			
ACCESSION			
BF398646			
VERSION			
BF398646.1 GI:11383643			
KEYWORDS			
EST.			
SOURCE			
Rattus norvegicus (Norway rat)			
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
Rattus.			
REFERENCE			
1 (bases 1 to 586)			
AUTHORS			
Bonaldo,M.F., Lennon,G. and Soares,M.B.			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
COMMENT			
Normalization and subtraction: two approaches to facilitate gene discovery			
Genome Res. 6 (9), 791-806 (1996)			
97044477			
889548			
Contact: Soares, MB			
Coordinated Laboratory for Computational Genomics			
University of Iowa			
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA			
Tel: 319 335 8250			
Fax: 319 335 9565			
Email: bento-soares@uiowa.edu			
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized embryo at 13 dpc library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)			
Seq primer: M13 Forward			
POLYA=Yes.			
FEATURES			
Location/Qualifiers			
1..586			
/organism="Rattus norvegicus"			
/mol_type="mRNA"			
/strain="Sprague-Dawley"			
/db_xref="taxon:10116"			
/clone="UI-R-BS2-ber-c-04-0-UI"			
/dev_stage="embryonic 13 dpc"			
/lab_host="DH10B (Life Technologies)"			
/clone_lib="UI-R-BS2"			
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BS2 library is a substracted library derived from 13 dpc whole embryo tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)			
TAG_ISSUE=embryo at 13 dpc			
TAG_LIB=UI-R-BS2			
TAG_SEQ=AATCC"			
ORIGIN			
Query Match			
Best Local Similarity 12.9%; Score 561; DB 10; Length 586;			
Matches 572; Conservative 0; Mismatches 0; Indels 1; Gaps 1;			
QY	3419	CTCTTCAAGACCCGGTCTAGCAAGTGACCACTATACTAATACCCCGGTCAACAAC	3478
Db	586	CTCTTCAAGACCCGGTCTAGCAAGTGACCACTATACTAATACCCCGGTCAACAAC	527
QY	3479	GTCATCCACAGGAAACCCCAATCAGTGTCAAGCAAGATGGTTCATCTCAGCGGCCTAC	3538
Db	526	GTCATCCACAGGAAACCCCAATCAGTGTCAAGCAAGAT-GGTTCATCTCAGCGGCCTAC	468
QY	3539	CCCCACCCGCATTCTCTATGTCAAAAGGTATGAAAGCTGGAAGCCAGTAGTGGCAGCCTC	3598
Db	467	CCCCACCCGCATTCTCTATGTCAAAAGGTATGAAAGCTGGAAGCCAGTAGTGGCAGCCTC	408
QY	3599	AGGAGCAGGAAATCTGACCAAAATTCAGCCTCGAGCTGAGACTCAGTCTATGAAAAATAGA	3658
Db	407	AGGAGCAGGAAATCTGACCAAAATTCAGCCTCGAGCTGAGACTCAGTCTATGAAAAATAGA	348
QY	3659	GCTGAAGAAATCTGCAGCAGCAGCAGTGCCTCTCTTTGGAGGGGGAGGGCTGAGGGCA	3718
Db	347	GCTGAAGAAATCTGCAGCAGCAGCAGTGCCTCTCTTTGGAGGGGGAGGGCTGAGGGCA	288
QY	3719	GTGGCTAAGGGGGTATGTTGTAAGGATGCTACTGCTGAGTGGAAACAAACCTTCTCTG	3778
Db	287	GTGGCTAAGGGGGTATGTTGTAAGGATGCTACTGCTGAGTGGAAACAAACCTTCTCTG	228

QY 3779 TGCCAACCCCTTCCCTTGACTACTAAATTAAGTTTAAATATCTTGTTTATAAAATAACC 3838
|||||
Db 227 TGCCAACCCCTTCCCTTGACTACTAAATTAAGTTTAAATATCTTGTTTATAAAATAACC 168
|||||
QY 3839 ATTTAATAGCCATGCACCCCTCCCATTTTGTGCATCTGTTTCAATGCAGGGGAATAGA 3898
|||||
Db 167 ATTTAATAGCCATGCACCCCTCCCATTTTGTGCATCTGTTTCAATGCAGGGGAATAGA 108
|||||
QY 3899 ATTAATTAGCAGAAATTTCTGTTTCTGTAAGTTCTGTTGAAGATGTTGGTCCAGTTCAGT 3958
|||||
Db 107 ATTAATTAGCAGAAATTTCTGTTTCTGTAAGTTCTGTTGAAGATGTTGGTCCAGTTCAGT 48
|||||
QY 3959 TTTACTTCTAGCATGTGGCCCATTCAGGTTAG 3991
|||||
Db 47 TTTACTTCTAGCATGTGGCCCATTCAGGTTAG 15
|||||

RESULT 14
AW525176/c
LOCUS
DEFINITION
UI-R-BJ0p-aio-f-08-0-UI.s1 UI-R-BJ0p Rattus norvegicus cDNA clone
UI-R-BJ0p-aio-f-08-0-UI 3', mRNA sequence.
ACCESSION
AW525176
VERSION
AW525176.1 GI:7167561
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 553)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL
MEDLINE
PUBMED
COMMENT
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized ventricle at 15 dpc library cDNA Library Preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source
1. .553
/location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ0p-aio-f-08-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-BJ0p"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ0p
library is a subtracted library derived from the UI-R-AA1,
UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and
UI-R-AG1 libraries. These libraries represent tissues from
rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
canal at 15 dpc, and ventricle at 13 dpc. The tag is a
string of 5-6 nucleotides present between the Not I site

and the oligo-dT track. The library was constructed as
described by Bonaldo, Lennon and Soares, Genome Research
6: 791-806, 1996.
TAG_TISSUE=ventricle at 15 dpc
TAG_LIB=UI-R-BJ0p
TAG_SEQ=GTGTC"

ORIGIN

Query Match 12.7%; Score 553; DB 10; Length 553;
Best Local Similarity 100.0%; Pred. No. 6.3e-73;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3799 TACTAATTTAAGTTTAAATATCTTGTTTATAAAATAACCATTTAATAGCCATGCACCCC 3858
|||||
Db 553 TACTAATTTAAGTTTAAATATCTTGTTTATAAAATAACCATTTAATAGCCATGCACCCC 494
|||||
QY 3859 CCTCCCATTTTGTGCATCTGTTTCAATGCAGGGGAATAGAATTAATTAGCAGAAATTTCTG 3918
|||||
Db 493 CCTCCCATTTTGTGCATCTGTTTCAATGCAGGGGAATAGAATTAATTAGCAGAAATTTCTG 434
|||||
QY 3919 TTTGCTGAATGTTCTGTTGAAGATGTTGGTCCAGTTCAGTTTTTACTTCTAGCATGTGGCC 3978
|||||
Db 433 TTTGCTGAATGTTCTGTTGAAGATGTTGGTCCAGTTCAGTTTTTACTTCTAGCATGTGGCC 374
|||||
QY 3979 CCATTCAAGGTAGCTCACGAGTTGTGAAGCCCTCAATATCGTCACCGAGAGATTGTGAGG 4038
|||||
Db 373 CCATTCAAGGTAGCTCACGAGTTGTGAAGCCCTCAATATCGTCACCGAGAGATTGTGAGG 314
|||||
QY 4039 ACCACATTACATATGCTCCCAAAGGCTGGCTCCCAATTTTCCCTAATTGTAAGCCAACTTT 4098
|||||
Db 313 ACCACATTACATATGCTCCCAAAGGCTGGCTCCCAATTTTCCCTAATTGTAAGCCAACTTT 254
|||||
QY 4099 AATAGACTCAGTTCTGTGATTTTTTTTTTCCAAAAAATAATATTTTGAATAGGACAGAG 4158
|||||
Db 253 AATAGACTCAGTTCTGTGATTTTTTTTTTCCAAAAAATAATATTTTGAATAGGACAGAG 194
|||||
QY 4159 TTTAACAGTTTGCATTTTGCACATCAAGCCATGAGTTTGATATATGGGTTATAAGAAAA 4218
|||||
Db 193 TTTAACAGTTTGCATTTTGCACATCAAGCCATGAGTTTGATATATGGGTTATAAGAAAA 134
|||||
QY 4219 GAATACTTTCAGAGCTATCACAGGGTCTCTAAACTTTTGGAAAAACAAGCCCTTAATA 4278
|||||
Db 133 GAATACTTTCAGAGCTATCACAGGGTCTCTAAACTTTTGGAAAAACAAGCCCTTAATA 74
|||||
QY 4279 TGACCTCAGGAAACAATTTGAACATGAAATAAAATGGAATGAACTGTGGAAATCTTAAAA 4338
|||||
Db 73 TGACCTCAGGAAACAATTTGAACATGAAATAAAATGGAATGAACTGTGGAAATCTTAAAA 14
|||||
QY 4339 AAAAAAAAAAAAAA 4351
|||||
Db 13 AAAAAAAAAAAAAA 1
|||||

RESULT 15
CB612875
LOCUS
DEFINITION
CB612875 540 bp mRNA linear EST 07-APR-2003
AMGNNUC:NRDGL-00033-F9-A nrdgl (10855) Rattus norvegicus cDNA clone
nrdgl-00033-f9 5', mRNA sequence.
ACCESSION
CB612875
VERSION
CB612875.1 GI:29572763
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 540)
Amgen EST Program.
TITLE
Amgen Rat EST Program
JOURNAL
Unpublished (2003)
COMMENT
Contact: Dan Fitzpatrick
Amgen, Inc
One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA

Tel: 805 447-4881	
Plate: 00033 row: f column: 9.	
FEATURES source	Location/Qualifiers
	1. .540
	/organism="Rattus norvegicus"
	/mol_type="mRNA"
	/db_xref="taxon:10116"
	/clone="nrdg1-00033-f9"
	/tissue_type="Dorsal Root Ganglia"
	/clone_lib="nrdg1 (10855)"
	/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; rat
	dorsal root ganglia"
ORIGIN	
Query Match 12.4%; Score 540; DB 14; Length 540;	
Best Local Similarity 100.0%; Pred. No. 5.6e-71;	
Matches 540; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	629 GAAGCACAAAGCAGCTGACTATACATGAACAAGAGCGACGACTTCACCAACCTGCTGGAGCA 688
Db	1 GAAGCACAAAGCAGCTGACTATACATGAACAAGAGCGACGACTTCACCAACCTGCTGGAGCA 60
Qy	689 GGAGCGAGAGAGGTTGAAAAAGCTCCTTGAACAAGAAAAAGCTTACCAAGCCCGCAAAGA 748
Db	61 GGAGCGAGAGAGGTTGAAAAAGCTCCTTGAACAAGAAAAAGCTTACCAAGCCCGCAAAGA 120
Qy	749 AAAGGAAAACGCTAAGCGGCTCAACAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCTT 808
Db	121 AAAGGAAAACGCTAAGCGGCTCAACAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCTT 180
Qy	809 CGCCCTCATGTTGGTGACGAGAGGCAGATGCACATCGAGCAACTGGGCTGCAGAGTCA 868
Db	181 CGCCCTCATGTTGGTGACGAGAGGCAGATGCACATCGAGCAACTGGGCTGCAGAGTCA 240
Qy	869 GAAAGTCCAGGACCTCACTCAGAAAGCTGAGGGAGGAGGAAGAAAAAAGCTCAAAAGCGGTCA 928
Db	241 GAAAGTCCAGGACCTCACTCAGAAAGCTGAGGGAGGAGGAAGAAAAAAGCTCAAAAGCGGTCA 300
Qy	929 TTACAAATCCAAGGAAGACCGCCAGAAAGCTGCTCAAGTTAGAAGTGGACTTCGAACACAA 988
Db	301 TTACAAATCCAAGGAAGACCGCCAGAAAGCTGCTCAAGTTAGAAGTGGACTTCGAACACAA 360
Qy	989 GGCCTCGAGGTTTCCCAAGGAGCACGAGAGATGAACGCCAAATTTGGCGAATCAAGAAATC 1048
Db	361 GGCCTCGAGGTTTCCCAAGGAGCACGAGAGATGAACGCCAAATTTGGCGAATCAAGAAATC 420
Qy	1049 TCACAACCGGCAACTTCGACTCAAACCTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGA 1108
Db	421 TCACAACCGGCAACTTCGACTCAAACCTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGA 480
Qy	1109 AGAGACCAATAAAAGCCTTCAGAAGCGCAGAGGAGAGCTCCAGGAGCTGAGAGAGAAAAT 1168
Db	481 AGAGACCAATAAAAGCCTTCAGAAGCGCAGAGGAGAGAGCTCCAGGAGCTGAGAGAGAAAAT 540

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein - protein search, using sw model

Run on: September 7, 2004, 12:59:38 ; Search time 150 Seconds
(without alignments)
2282.984 Million cell updates/sec

Title: US-10-788-793-2
Perfect score: 6082
Sequence: 1 MRSRNQGGESSNGHVSCP.....KIELKKSAASTASLGGKG 1212

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6082	100.0	1212	6	ABP97029	Abp97029 Rat L-FIL
2	5696.5	93.7	1213	4	AAM40016	Aam40016 Human pol
3	5696.5	93.7	1213	6	ABP97031	Abp97031 Human L-F
4	4838	79.5	965	6	ABP97030	Abp97030 Rat S-FIL
5	3051	50.2	653	4	AAB93085	Aab93085 Human pro
6	2586	42.5	551	4	AAB93244	Aab93244 Human pro
7	2452.5	40.3	1133	6	ABR62252	AbR62252 GPBP-inte
8	2444.5	40.2	1133	6	ABR62253	AbR62253 GPBP-inte
9	2444.5	40.2	1135	6	ABR62251	AbR62251 GPBP-inte
10	1926	31.7	893	6	ABR69651	AbR69651 Human CGD
11	1828	30.1	764	6	ABR62250	AbR62250 GPBP-inte
12	1675	27.5	785	6	ABR62257	AbR62257 GPBP-inte
13	1645.5	27.1	752	5	AAU84329	Aau84329 Protein D
14	1645.5	27.1	752	6	ABR92070	AbR92070 Human cer
15	1040.5	17.1	233	4	AAM41802	Aam41802 Human pro
16	929	15.3	365	4	AAB63771	Aab63771 Human pro
17	893	14.7	1053	6	ABR64181	AbR64181 Angiogene
18	893	14.7	1076	7	ADC31075	Adc31075 Human nov
19	893	14.7	1087	6	ABG74689	Abg74689 Human CGD
20	780.5	12.8	439	3	AAB58157	Aab58157 Lung canc
21	769.5	12.7	350	6	ABR62249	AbR62249 GPBP-inte
22	677	11.1	265	6	ABR62248	AbR62248 GPBP-inte
23	618.5	10.2	384	6	ABR62264	AbR62264 GPBP-inte
24	618.5	10.2	386	6	ABR62256	AbR62256 GPBP-inte
25	602	9.9	274	7	ADB65199	Adb65199 Human pro

26	569	9.4	190	4	AAB63826	Aab63826 Human pro
27	526.5	8.7	240	6	ABR62263	AbR62263 GPBP-inte
28	518.5	8.5	240	6	ABR62247	AbR62247 GPBP-inte
29	512	8.4	225	4	AAB63875	Aab63875 Human pro
30	476	7.8	2349	6	AAO16359	Aao16359 Human tra
31	476	7.8	2415	4	ABG20279	Abg20279 Novel hum
32	474	7.8	2400	4	ABG20278	Abg20278 Novel hum
33	469	7.7	1960	4	AAM78854	Aam78854 Human pro
34	469	7.7	1960	7	ADB70370	Adb70370 PAC 6802
35	469	7.7	1960	7	ADD47946	Add47946 Human Pro
36	469	7.7	2143	4	ABG01716	Abg01716 Novel hum
37	469	7.7	2688	4	AAM40883	Aam40883 Human pol
38	468.5	7.7	1959	5	ABJ10604	Abj10604 Human nov
39	468	7.7	2663	4	AAM39097	Aam39097 Human pol
40	464.5	7.6	1790	6	ABR53116	AbR53116 Protein s
41	463.5	7.6	1963	4	AAM79838	Aam79838 Human pro
42	461.5	7.6	1881	5	ABP73809	Abp73809 Candida a
43	459	7.5	2633	4	ABG06505	Abg06505 Novel hum
44	452	7.4	1857	5	AAU84350	Aau84350 Protein M
45	452	7.4	1938	6	ABR92127	AbR92127 Human cer

ALIGNMENTS

RESULT 1
ABP97029
ID ABP97029 standard; protein; 1212 AA.
XX

AC ABP97029;

XX
DT 18-JUN-2003 (first entry)

XX
DE Rat L-FILIP protein SEQ ID NO:2.

XX
KW L-FILIP; S-FILIP; filamin-interacting protein; cell migration;

KW cell death; cytosstatic; neuroprotective; immunosuppressive; cancer;

KW tumour metastasis; transplantation therapy.

XX
OS Rattus norvegicus.

XX
PN WO2003018804-A1.

XX
PD 06-MAR-2003.

PF 29-JUL-2002; 2002WO-JP007676.

XX
PR 27-AUG-2001; 2001JP-00256910.

XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.

XX
PI Sato M, Nagano T;

XX
DR WPI; 2003-268423/26.

DR
N-PSDB; ACC45354.

XX
PT Proteins controlling cell migration and cell death and their encoded

PT DNAs, applicable in developing drugs for treating or suppressing cancer

PT or tumor metastasis or as regulators of cell migration for

PT transplantation.

XX
PS Claim 1; Page 54-60; 96pp; Japanese.

XX
CC The present sequence represents rat L-FILIP which is a filamin-

CC interacting protein. L-FILIP has a function of controlling cell migration

CC and cell death. L-Filip has cytosstatic, neuroprotective and

CC immunosuppressive activities. The L-FILIP protein can be used for

CC controlling cell migration and cell death, which is applicable in

CC developing drugs for treating or suppressing cancer or tumour metastasis

CC or as regulators of cell migration for transplantation therapy, and also

CC for controlling the mobility and cell death of nerve cells, promoting

CC decomposition of the actin-binding protein e.g. filamin-interacting

CC protein in the treatment of preiventricular nodular heterotopia

XX	SQ	Sequence	1212	AA;
		Query Match	100.0%;	Score 6082; DB 6; Length 1212;
		Best Local Similarity	100.0%;	Pred. No. 1.2e-303;
		Matches 1212; Conservative	0; Mismatches	0; Indels 0; Gaps 0;
QY	1	MRSRNGGESSNGHVSCPKSSIISSDGGKGPSEDAKKNKANRKEEDVMASGTIKRHLKP	60	
DB	1	MRSRNGGESSNGHVSCPKSSIISSDGGKGPSEDAKKNKANRKEEDVMASGTIKRHLKP	60	
QY	61	SGSEKTKKSVELSKEDLIQLLSIMEGELQAREDVIHMLRTEKTKPEVLEAHVGSAPPE	120	
DB	61	SGSEKTKKSVELSKEDLIQLLSIMEGELQAREDVIHMLRTEKTKPEVLEAHVGSAPPE	120	
QY	121	KVLRLVLRDAIIAQEKSTIGEDVYEKPISELDRLEEKOKETVYRMLLEQLLLAEKCHRRTVY	180	
DB	121	KVLRLVLRDAIIAQEKSTIGEDVYEKPISELDRLEEKOKETVYRMLLEQLLLAEKCHRRTVY	180	
QY	181	ELENEKHKHTDYMNKSDFTNLLEQERERLKKLLEQEKAYQARKEKENAKRLNKLRDELV	240	
DB	181	ELENEKHKHTDYMNKSDFTNLLEQERERLKKLLEQEKAYQARKEKENAKRLNKLRDELV	240	
QY	241	KLKSFALMLVDEROMHITEQLGLSQKVQDLTKQLREEEKCLKAVTYKSKEDRKLLKLEV	300	
DB	241	KLKSFALMLVDEROMHITEQLGLSQKVQDLTKQLREEEKCLKAVTYKSKEDRKLLKLEV	300	
QY	301	DFEHKASRFSQEHHEFNMAKLANQESHNRQLRLKLVGLSQRIEELEETNKSQKABEELQE	360	
DB	301	DFEHKASRFSQEHHEFNMAKLANQESHNRQLRLKLVGLSQRIEELEETNKSQKABEELQE	360	
QY	361	LREKIAKGECGNSSLMABVESLRKRVLEMEGKDEEITKTEAQCRELKKKQLEEEHHSKEL	420	
DB	361	LREKIAKGECGNSSLMABVESLRKRVLEMEGKDEEITKTEAQCRELKKKQLEEEHHSKEL	420	
QY	421	RLEVEKLQKRMSELEKLEBAFSPRSKSECTQLHLNLEKEKNLT KDLLNELEVVKSRVKELE	480	
DB	421	RLEVEKLQKRMSELEKLEBAFSPRSKSECTQLHLNLEKEKNLT KDLLNELEVVKSRVKELE	480	
QY	481	CSESRLKAEKLSLKDITKLKSFVTMLVDERKNMMEKIKQEERKVDGLNKNFKVEQGVKM	540	
DB	481	CSESRLKAEKLSLKDITKLKSFVTMLVDERKNMMEKIKQEERKVDGLNKNFKVEQGVKM	540	
QY	541	DVTEKLI EESKLLKLKSEMEEEKYSLTKERDELGMGLRSEERSCELSVDDLKRRLD	600	
DB	541	DVTEKLI EESKLLKLKSEMEEEKYSLTKERDELGMGLRSEERSCELSVDDLKRRLD	600	
QY	601	GIEEVEIREINGRSCKGSEFTCPEDNKIRELTLEIERLKKRLOQLEVVEGDLMKTEDEYD	660	
DB	601	GIEEVEIREINGRSCKGSEFTCPEDNKIRELTLEIERLKKRLOQLEVVEGDLMKTEDEYD	660	
QY	661	QLEQKFRTEQDKANFLSQOLEEIKHQMAKHKAIEKGEAVSQEAELRHRFRLEEAKSRDLQ	720	
DB	661	QLEQKFRTEQDKANFLSQOLEEIKHQMAKHKAIEKGEAVSQEAELRHRFRLEEAKSRDLQ	720	
QY	721	AEVQALKEKIHLMNKEQDLSQLQVDYSVLQORFMEETKNKNMGREVNLTKLELSKR	780	
DB	721	AEVQALKEKIHLMNKEQDLSQLQVDYSVLQORFMEETKNKNMGREVNLTKLELSKR	780	
QY	781	YSRALRPSGNGRRMVDVPVASTGVQTEAVCGDAAEEETPAVFIRKSFQENHIMSRLQV	840	
DB	781	YSRALRPSGNGRRMVDVPVASTGVQTEAVCGDAAEEETPAVFIRKSFQENHIMSRLQV	840	
QY	841	GLKKPMERSVLD RYPPAANELTMRKSWIPWMRKRENGPSTPQEKGRPNQAGHPGELV	900	
DB	841	GLKKPMERSVLD RYPPAANELTMRKSWIPWMRKRENGPSTPQEKGRPNQAGHPGELV	900	
QY	901	LAPKQGPLHVRVTPDHENSTATLEITSPSTSEFFSSTTVIPTLGNQKPRITIIIPSPNVM	960	
DB	901	LAPKQGPLHVRVTPDHENSTATLEITSPSTSEFFSSTTVIPTLGNQKPRITIIIPSPNVM	960	
QY	961	SQPKSADPTLGPERRAMPVITITISREKSPGGRSAFADRPASPIQIMTVSTSAATEI	1020	

Db	961	SQPKSADPTLGP	RAMSPVTITTTISREKSP	EGGSAFADRPASPIQIMTVSTSAAPTEI	1021
Qy	1021	AVSPSEQEVP	MGRTILKVTPEKQTVP	PAPVRKYNSNANIITTTEDNKIHLG	1080
Db	1021	AVSPSEQEVP	MGRTILKVTPEKQTVP	PAPVRKYNSNANIITTTEDNKIHLG	1080
Qy	1081	AAEGVSPVITVR	PVNVTAKEVSTGT	VLRSRPNHLSSRP	1140
Db	1081	AAEGVSPVITVR	PVNVTAKEVSTGT	VLRSRPNHLSSRP	1140
Qy	1141	TQSVSGDQSS	QRPPTTRIPMSKGMKAGKPV	VAAAGAGNLT	1200
Db	1141	TQSVSGDQSS	QRPPTTRIPMSKGMKAGKPV	VAAAGAGNLT	1200
Qy	1201	ASSTASLG	GKG 1212		
Db	1201	ASSTASLG	GKG 1212		
RESULT 2					
AAM40016					
ID	AAM40016	standard; protein; 1213 AA.			
XX	AC	AAM40016;			
XX	22-OCT-2001	(first entry)			
DT					
XX		Human polypeptide SEQ ID NO 3161.			
XX		Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;			
KW		peripheral nervous system; neuropathy; central nervous system; CNS;			
KW		Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;			
KW		amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;			
KW		chemokinetic; thrombolytic; drug screening; arthritis; inflammation;			
KW		leukaemia.			
XX					
OS		Homo sapiens.			
XX					
PN		WO200153312-A1.			
XX					
PD		26-JUL-2001.			
XX					
PF		26-DEC-2000; 2000WO-US034263.			
XX					
PR		23-DEC-1999; 99US-00471275.			
PR		21-JAN-2000; 2000US-00488725.			
PR		25-APR-2000; 2000US-0052317.			
PR		20-JUN-2000; 2000US-00598042.			
PR		19-JUL-2000; 2000US-00620312.			
PR		03-AUG-2000; 2000US-00653450.			
PR		14-SEP-2000; 2000US-00662191.			
PR		19-OCT-2000; 2000US-00693036.			
PR		29-NOV-2000; 2000US-00727344.			
XX					
XX		(HYSE-) HYSEQ INC.			
XX					
PI		Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;			
PI		Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;			
PI		Zhou P, Goodrich R, Drmanac RT;			
XX					
DR		WPI; 2001-442253/47.			
DR		N-PSDB; AAI59172.			
XX					
PT		Novel nucleic acids and polypeptides, useful for treating disorders such			
PT		as central nervous system injuries.			
XX					
PS		Example 4; SEQ ID NO 3161; 10078pp; English.			
XX					
CC		The invention relates to human nucleic acids (AAI57798-AAI61369) and the			
CC		encoded polypeptides (AAM38642-AAM42213) with nootropic,			
CC		immunosuppressant and cytostatic activity. The polynucleotides are useful			
CC		in gene therapy. A composition containing a polypeptide or polynucleotide			
CC		of the invention may be used to treat diseases of the peripheral nervous			

CC or as regulators of cell migration for transplantation therapy, and also
CC for controlling the mobility and cell death of nerve cells, promoting
CC decomposition of the actin-binding protein e.g. filamin-interacting
CC protein in the treatment of preinvenitrlcular nodular heterotopia
xx
SQ Sequence 1213 AA;

Query Match 93.7%; Score 5696.5; DB 6; Length 1213;
Best Local Similarity 93.5%; Pred. No. 7.1e-284;
Matches 1134; Conservative 37; Mismatches 41; Indels 1; Gaps 1;

Qy 1 MRSRNOGGESSNGHVSCPKSSIISSDGGKGPSEDA-KKNKANRKEEDVNASGTIKRHUK 59
Db 1 MRSRNOGGEASDGHISCPKPSIIIGNAGEKSLSEDAKKKKSNRKEDDVNASGTVKRHLK 60

Qy 60 PSGESEKTKKSVELSKEDLIQLLSIMEGELQAREDVIHMLRTEKTKPEVLEAHYGSAP 119
Db 61 TSGECERTKKSLSESKEDLIQLLSIMEGELQAREDVIHMLKTEKTKPEVLEAHYGSAP 120

Qy 120 EKVLRLVLRDAILAQEKSGEDVYEKPISELDRLEEKQKETYRRMLEQLLLAEKCHRRTV 179
Db 121 EKVLRLVLRDAILAQEKSGEDVYEKPISELDRLEEKQKETYRRMLEQLLLAEKCHRRTV 180

Qy 180 YELENEKHKHTDYMNKSDFTNLLEQERERLKLLEQEKAYQARKEKENAKRLNKLRLDEL 239
Db 181 YELENEKHKHTDYMNKSDFTNLLEQERERLKLLEQEKAYQARKEKENAKRLNKLRLDEL 240

Qy 240 VKLKSFALMLVDERQMHIETQLGLOSQKVQDLTQKLREEEBEKLKAVTYKSKEDRQKLLKLE 299
Db 241 VKLKSFALMLVDERQMHIETQLGLOSQKVQDLTQKLREEEBEKLKAITSKSKEDRQKLLKLE 300

Qy 300 VDFEHKASRFSQHEHEMNAKLANQESHNRQLRLKLVGLSORIEELEETNKSLOKAEELQ 359
Db 301 VDFEHKASRFSQHEHEMNAKLANQESHNRQLRLKLVGLTQRIELEEETNKNLQKAEELQ 360

Qy 360 ELREKIAKGECCGNSSLMAEVESLRKRVLEMEGKDEEITKTEAQCRELKKLOEEHHKSKE 419
Db 361 ELRDKIAKGECCGNSSLMAEVENLRKRVLEMEGKDEEITKTESQCRELRKKLOEEHHKSKE 420

Qy 420 LRLEVEKLOKRMSELEKLEEAFFSRKSECTQLHLNLEKEKNLT KDLLNLELVVKS RVKEL 479
Db 421 LRLEVEKLOKRMSELEKLEEAFFSKSECTQLHLNLEKEKNLT KDLLNLELVVKS RVKEL 480

Qy 480 ECSESRLKAEKSLKDDLT KLKSFVTVMLVDERKNMMEKIKQEEKVVDGLNKNFKVEQGKV 539
Db 481 ECSESRLKAEKSLKDDLT KLKSFVTVMLVDERKNMMEKIKQEEKVVDGLNKNFKVEQGKV 540

Qy 540 MDVTEKLIIESKLLKLKSEMEEKEYSLTKERDELGMKLRSEERSCELSVDDLKKRL 599
Db 541 MDVTEKLIIESKLLKLKSEMEEKVYNLTREDELIGLKSEEEKSSELSVDDLKKRL 600

Qy 600 DGIEEVEREINRGRCKGSEFTCPEDNKKIRELTLEIERLKKRLQQLLEVVEGDLMKTEDEY 659
Db 601 DGIEEVEREITRGRSRKSELTCPEDNKKIKELTLEIERLKKRLQQLLEVVEGDLMKTEDEY 660

Qy 660 DQLEQKFRTEQDKANFLSQOLEEIKHQMAKHAIEKGEAVSQEAELRHRFRLEEAKSRL 719
Db 661 DQLEQKFRTEQDKANFLSQOLEEIKHQIAKNKAIEKGEVVSQEAELRHRFRLEEAKSRL 720

Qy 720 QAEVQALKEKIHLMNKEPQLSQLDVSVLQQRFEEMEEETKNMGMREVLNLTKELELSK 779
Db 721 KAEVQALKEKIHLMNKEPQLSQLDVSVLQQRFMEEENKNMGMQEVNLTKELELSK 780

Qy 780 RYSRALRPSGNRRMVDVPVASTGVQTEAVCGDAEEETPAVFIRKSFQENHIMSNLRQ 839
Db 781 RYSRALRPSVNGRRMVDVPVTSTGVQTDVAVGEAAEEETPAVFIRKSFQENHIMSNLRQ 840

Qy 840 VGLKKPNERSSVLD RYPPAANELTMRKSWIPWMRKRENGPSTPOEKGPPNQAGHPGEL 899
Db 841 VGLKKPVERSSVLD RYPPAANELTMRKSWIPWMRKRENGPSTPOEKGPRTNSSPGHPGEV 900

Qy 900 VLAPKQGOPLHIRVTPDHENSTATLEITSPTSEEFFSSTVIPTLGNQKPRITIIPSPNV 959
||:|||||

Db 901 VLSPKQGOPLHIRVTPDHENSTATLEITSPTSEEFFSSTVIPTLGNQKPRITIIPSPNV 960

Qy 960 MSQPKKSADPTLGPERAMSPVTITTSREKSPGEGRSFAFADRPASPIQIMTVSTSAAPTE 1019
Db 961 MPQKQKSGDTTLGPERAMSPVTITTFREKTPESGRGAFADRPSTPIQIMTVSTSAAPAE 1020

Qy 1020 IAVSPESQEVPMGRTILKVTPEKQTVPAVRKYNNSNANITTEDNKIHIHLGSOFKRSPG 1079
Db 1021 IAVSPESQEMPGRITILKVTPEKQTVPTVRKYNNSNANITTEDNKIHIHLGSOFKRSPG 1080

Qy 1080 PAAEGVSPVITVRPVNVTAKEVSTGTVLRSRPNHLSSRPGASKVTSTITITPVTTSSSTR 1139
Db 1081 TSGEGVSPVITVRPVNVTAKEVSTGTVLRSRPNHLSSRPGASKVTSTITITPVTTSSAR 1140

Qy 1140 GTQSVSGQDSSQRPTPTRIIPMSKGMKAGKPVVAASGAGNLTKFPRAETQSMKIELKKS 1199
Db 1141 GTQSVSGQDSSQRPTPTRIIPMSKGMKAGKPVVAAPGAGNLTKFEPRAETQSMKIELKKS 1200

Qy 1200 AASSTASLGGGKG 1212
Db 1201 AASSTSLGGGKG 1213

RESULT 4
ABP97030
ID ABP97030 standard; protein; 965 AA.
XX
AC ABP97030;
XX
DT 18-JUN-2003 (first entry)
XX
DE Rat S-FILIP protein SEQ ID NO:4.
XX
KW L-FILIP; S-FILIP; filamin-interacting protein; cell migration;
KW cell death; cytostatic; neuroprotective; immunosuppressive; cancer;
KW tumour metastasis; transplantation therapy.
XX
OS Rattus norvegicus.
XX
PN WO2003018804-A1.
XX
PD 06-MAR-2003.
XX
PF 29-JUL-2002; 2002WO-JP007676.
XX
PR 27-AUG-2001; 2001JP-00256910.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI Sato M, Nagano T;
XX
DR WPI; 2003-268423/26.
DR N-PSDB; ACC45355.
XX
PT Proteins controlling cell migration and cell death and their encoded
PT DNAs, applicable in developing drugs for treating or suppressing cancer
PT or tumor metastasis or as regulators of cell migration for
PT transplantation.
XX
PS Claim 4; Page 68-73; 96pp; Japanese.
XX
CC The present sequence represents rat S-FILIP which is a filamin-
CC interacting protein. S-FILIP has a function of controlling cell migration
CC and cell death. S-FILIP has cytostatic, neuroprotective and
CC immunosuppressive activities. The S-FILIP protein can be used for
CC controlling cell migration and cell death, which is applicable in
CC developing drugs for treating or suppressing cancer or tumour metastasis
CC or as regulators of cell migration for transplantation therapy, and also
CC for controlling the mobility and cell death of nerve cells, promoting
CC decomposition of the actin-binding protein e.g. filamin-interacting
CC protein in the treatment of preinvenitrlcular nodular heterotopia
XX
SQ Sequence 965 AA;

Query Match		79.5%;	Score 4838;	DB 6;	Length 965;			
Best Local Similarity		100.0%;	Pred. No. 6.1e-240;					
Matches 965;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;			
Qy	248	MLVDERQMHI	EQGLQSQVQDLTQKLR	EEEEKLVAVTYKSKEDRQKLLK	LEVDFEHKAS 307			
Db	1	MLVDERQMHI	EQGLQSQVQDLTQKLR	EEEEKLVAVTYKSKEDRQKLLK	LEVDFEHKAS 60			
Qy	308	RFSQHEEMNA	KLANQESHNRQLRLKLVGLS	QRIIELEETNKSLOKAEELQEL	REKIAK 367			
Db	61	RFSQHEEMNA	KLANQESHNRQLRLKLVGLS	QRIIELEETNKSLOKAEELQEL	REKIAK 120			
Qy	368	GECCNSSLMA	EVESLRKRVLEMEGKDDEEITK	TEAQCRELKKKLOEEHHSKELR	LEVVKL 427			
Db	121	GECCNSSLMA	EVESLRKRVLEMEGKDDEEITK	TEAQCRELKKKLOEEHHSKELR	LEVVKL 180			
Qy	428	QKRMELEKLE	EAFRSKSECTQLHLNLEKEKNLT	KDLLNELEVVKSRVKELECS	SRLE 487			
Db	181	QKRMELEKLE	EAFRSKSECTQLHLNLEKEKNLT	KDLLNELEVVKSRVKELECS	SRLE 240			
Qy	488	KAEKSLKDDL	TKLKSFTVMLVDERKNMMEKIK	QOERKVDGLNKNFKVEQGVMDV	TEKLI 547			
Db	241	KAEKSLKDDL	TKLKSFTVMLVDERKNMMEKIK	QOERKVDGLNKNFKVEQGVMDV	TEKLI 300			
Qy	548	EESKLLKLK	SEMEEKEYSLTKERDEL	MGLRSEEEERSCELSVDLLK	KRLDGLIEEVER 607			
Db	301	EESKLLKLK	SEMEEKEYSLTKERDEL	MGLRSEEEERSCELSVDLLK	KRLDGLIEEVER 360			
Qy	608	EINRGSRCKG	SEFTCPEDNKIRELTLEIERL	KKRLQQLLEVVEGDLMKTE	DEYDQLEQKFR 667			
Db	361	EINRGSRCKG	SEFTCPEDNKIRELTLEIERL	KKRLQQLLEVVEGDLMKTE	DEYDQLEQKFR 420			
Qy	668	TEQDKANFLS	QQLEEKHOMAKHAKAIEKGE	AVSQEAEELHFRLEEA	KSRDLQAEVQALK 727			
Db	421	TEQDKANFLS	QQLEEKHOMAKHAKAIEKGE	AVSQEAEELHFRLEEA	KSRDLQAEVQALK 480			
Qy	728	EKIHEL	MNKEDQLSQLQVDSVQLQOR	FMEETKKNMGREVLNLTKE	LELSKRYSRALRP 787			
Db	481	EKIHEL	MNKEDQLSQLQVDSVQLQOR	FMEETKKNMGREVLNLTKE	LELSKRYSRALRP 540			
Qy	788	SGNGRRMVD	VPVASTGVQTEAVCGDA	AEETPAVFIRKSFQENH	IMSNLRQVGLKPKME 847			
Db	541	SGNGRRMVD	VPVASTGVQTEAVCGDA	AEETPAVFIRKSFQENH	IMSNLRQVGLKPKME 600			
Qy	848	RSSVLD	RYPPAANELTMRKSWIP	PWMKRKRENGPSTPQEK	GRPNQAGHPGELVLPKQGG 907			
Db	601	RSSVLD	RYPPAANELTMRKSWIP	PWMKRKRENGPSTPQEK	GRPNQAGHPGELVLPKQGG 660			
Qy	908	PLHIRV	TPDHENSTATLEIT	TSPTSEEFFSSTTVIPT	LGKQKPRITIIIPSPNVM	SQKPKSA 967		
Db	661	PLHIRV	TPDHENSTATLEIT	TSPTSEEFFSSTTVIPT	LGKQKPRITIIIPSPNVM	SQKPKSA 720		
Qy	968	DPTLGP	RAMSPVTITISREKSP	EGGRSAFADRPASPIQ	IMTVSTSAAPT	IAVSPESQ 1027		
Db	721	DPTLGP	RAMSPVTITISREKSP	EGGRSAFADRPASPIQ	IMTVSTSAAPT	IAVSPESQ 780		
Qy	1028	EVP	MGRITILKVTPEKQTV	PAPVRKYNSNANIIT	TEDNKIHILG	SQFKRSPGPA	AEVSP 1087	
Db	781	EVP	MGRITILKVTPEKQTV	PAPVRKYNSNANIIT	TEDNKIHILG	SQFKRSPGPA	AEVSP 840	
Qy	1088	VITVRP	VNVTAEKEVSTGTV	LRSPRNLSSRPGASK	VTSTITITPVT	TSSTRGTQ	SVSGQ 1147	
Db	841	VITVRP	VNVTAEKEVSTGTV	LRSPRNLSSRPGASK	VTSTITITPVT	TSSTRGTQ	SVSGQ 900	
Qy	1148	DGSSQR	PTTRIPMSKGMKAGK	PPVAAASGAGNLT	TKFOPRAETQ	SMKIELK	KSAA	STASL 1207
Db	901	DGSSQR	PTTRIPMSKGMKAGK	PPVAAASGAGNLT	TKFOPRAETQ	SMKIELK	KSAA	STASL 960
Qy	1208	GGGKG	1212					
Db	961	GGGKG	965					

RESULT 5	
AAB93085	
ID	AAB93085 standard; protein; 653 AA.
XX	
AC	AAB93085;
XX	
DT	26-JUN-2001 (first entry)
XX	
DE	Human protein sequence SEQ ID NO:11924.
XX	
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX	
OS	Homo sapiens.
XX	
PN	EP1074617-A2.
XX	
PD	07-FEB-2001.
XX	
PF	28-JUL-2000; 2000EP-00116126.
XX	
PR	29-JUL-1999; 99JP-00248036.
PR	27-AUG-1999; 99JP-00300253.
PR	11-JAN-2000; 2000JP-00118776.
PR	02-MAY-2000; 2000JP-00183767.
PR	09-JUN-2000; 2000JP-00241899.
XX	
PA	(HELI-) HELIX RES INST.
XX	
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX	
DR	WPI; 2001-318749/34.
XX	
PT	Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
PS	Claim 8; SEQ ID NO 11924; 2537pp + Sequence Listing; English.
XX	
CC	The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
SQ	Sequence 653 AA;

Query Match		50.2%;	Score 3051;	DB 4;	Length 653;
Best Local Similarity		92.3%;	Pred. No. 2e-148;		
Matches 603;		Conservative 19;	Mismatches 31;	Indels 0;	Gaps 0;
Qy	560	MEEKEYSLTKER	DELMGKLRSEERSCELS	CSVDLLKKRLD	GIEEVEREINRGSRCKGSE 619
Db	1	MEEKVYNLT	RRERDELIGLKGSEES	SELSVDLLKKRLD	GIEEVEREITGRSRCKGSE 60

QY 620 FTCPEDNKKIRELTLEIERLKKRLQOLEVVEGDLMKTEDEYDQLEQFRTEQDKANFLSQ 679
Db 61 LTCPEDNKKIKELTLEIERLKKRLQOLEVVEGDLMKTEDEYDQLEQFRTEQDKANFLSQ 120
QY 680 LEEIKHQMAKHKAIEKGEAVSQEAEILRHRFRLEEAKSRDLQAEVQALKKEIKIHELMNKEDQ 739
Db 121 LEEIKHQIAKNKAIEKGEVVSQEAEILRHRFRLEEAKSRDLKAEVQALKKEIKIHELMNKEDQ 180
QY 740 LSQLOVDYSVLQORPMEBEETKNKMGREVLNLTKELELSKRYSRALRPSGNGRRMVDVPV 799
Db 181 LSQLOVDYSVLQORPMEBEENKKNMGQEVNLTKELELSKRYSRALRPSVNGRRMVDVPV 240
QY 800 ASTGVQTEAVCGDAAEETPAVFIRKSFQEENHIMSNLRQVGLKKPMESSSVLDRYPPAA 859
Db 241 TSTGVQTDVAVSGEAAEETPAVFIRKSFQEENHIMSNLRQVGLKKPVERSSVLDRYPPAA 300
QY 860 NELTMRKSWIPWNRKRENGPSTPOBKGP RPNOGAGHPGELVLAPKQGQPLHIRVTPDHEN 919
Db 301 NELTMRKSWIPWNRKRENGPSTQEKGPRTNSSPGHPGEVVLSPKQGQPLHIRVTPDHEN 360
QY 920 STATLEITSPTSEFFFSSTTVITLGNQKPRITIIPSPNVMQPKKSADPTLGP RAMSP 979
Db 361 STATLEITSPTSEFFFSSTTVITLGNQKPRITIIPSPNVMQPKQKSGDPTLGP RAMSP 420
QY 980 VTITTSIREKSPGGRSAFADRPASPIQIMTVSTSAAPTETIAVSPESQEVPMGRTILKVT 1039
Db 421 VTITTSREKTPESGRGAFADRTSP IQIMTVSTSAAPAEIAVSPESQEMPMPGRTILKVT 480
QY 1040 PEKQTVPAVRKYNSNANIITTEDNKIHIHLGSGQFKRSPGPAEGVSPVITVRPVNVTAE 1099
Db 481 PEKQTVPTPVRRKYNSNANIITTEDNKIHIHLGSGQFKRSPGTSGEVSPVITVRPVNVTAE 540
QY 1100 KEVSTGTVLRSPRNHLSSRPGASKVTSTITITPVTTSSTRGTQSVSGQDSSQRPPTPRI 1159
Db 541 KEVSTGTVLRSPRNHLSSRPGASKVTSTITITPVTTSSTRGTQSVSGQDSSQRPPTPRI 600
QY 1160 PMSKGMKAGKPVVAASGAGNLTQFPRAETQSMKIELKKSAAASSTASLGSGKG 1212
Db 601 PMSKGMKAGKPVVAAPGAGNLTQFPRAETQSMKIELKKSAAASSTSLGSGKG 653
RESULT 6
AAB93244
ID AAB93244 standard; protein; 551 AA.
XX
AC AAB93244;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12253.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX

DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
PT
PT
XX
PS Claim 8; SEQ ID NO 12253; 2537bp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
XX
SQ Sequence 551 AA;
Query Match 42.5%; Score 2586; DB 4; Length 551;
Best Local Similarity 94.7%; Pred. No. 1.2e-124;
Matches 521; Conservative 19; Mismatches 10; Indels 0; Gaps 0;
QY 248 MLVDERQWHIEQLGLOQKQVQDLTQKLREEEKKLVATYKSKEDRQKLLKLEVD FEHKAS 307
Db 1 MLVDERQWHIEQLGLOQKQVQDLTQKLREEEKKLKAITSKSKEDRQKLLKLEVD FEHKAS 60
QY 308 RFSQHEEMNAKLANQESHNRQLRLKLVGLSQRIEEL EETNKSQKAEELQELREKIAK 367
Db 61 RFSQHEEMNAKLANQESHNRQLRLKLVGLTQRIEEL EETNKNLQKAEELQELRDKIAK 120
QY 368 GECGNSILMAEVESLRKRVLEMEGKDEEITKTEAQC RELKKKLOEEHHSKELRLEVEKL 427
Db 121 GECGNSILMAEVENLRKRVLEMEGKDEEITKTESQ RELRKKLQEEHHSKELRLEVEKL 180
QY 428 QKRMSELEKLEEA FSRKSECTQLHLNLEKEKNLT KDLLNELEVVKSRVKELECSERLE 487
Db 181 QKRMSELEKLEEA FSRKSECTQLHLNLEKEKNLT KDLLNELEVVKSRVKELECSERLE 240
QY 488 KAELS LKDDLT K LK SFTV MLVDERKNMMEKIKQE ERKVDGLNKNFKVEQGKVM DVTEKLI 547
Db 241 KAELS LKDDLT K LK SFTV MLVDERKNMMEKIKQE ERKVDGLNKNFKVEQGKVM DVTEKLI 300
QY 548 EESKLLKLKSEMERKEYSLTKERDEIMGKL RSEERSCELSCSV D L L K K R L D G I E E V E R 607
Db 301 EESKLLKLKSEMERKEYSLTKERDEIMGKL RSEERSCELSCSV D L L K K R L D G I E E V E R 360
QY 608 EINGRSCKGSEFTCPEDNKIRELTLEIERLKKRLQOLEVVEGDLMKTEDEYDQLEQKFR 667
Db 361 EITRGRSRKGS ELT C PEDNKI K E L T L E I E R L K K R L Q O L E V V E G D L M K T E D E Y D Q L E Q K F R 420
QY 668 TEQDKANFLSQLEEKHQMAKHKAIEKGEAVSQEAEILRHRFRLEEAKSRDLQAEVQALK 727
Db 421 TEQDKANFLSQLEEKHQIAKNKAIEKGEVVSQEAEILRHRFRLEEAKSRDLKAGVQALK 480
QY 728 EKIHLMNKEDQLSQLOVDYSVLQORFMEETKNKMGREVLNLTKELELSKRYSRALRPP 787
Db 481 EKIHLMNKEDQLSQLOVDYSVLQORFMEETKNKMGREVLNLTKELELSKRYSRALRPP 540

QY	788	SGNGRRMVDV	797	QY	116	SAPEKVLRLHRDA	LAQEKSIGEDVYEKPISELDRL	EEKQKETYRRMLEQLLLAEKCH	175	
Db	541	SVNGKRMVDV	550	Db	109	FVTPKKVLEALQ	RDAFAQAKSTPWQEDIYEKPMN	ELDQVVEKHKESYRRILGQLLVAEKSH	168	
RESULT 7										
ABR62252										
ID	ABR62252	standard; protein; 1133 AA.								
XX	AC	ABR62252;								
XX	DT	22-SEP-2003	(first entry)	176	RRTVYELENEKH	KHTDYMKNKSDFTNNLLEQERER	LKJLLEQEKAYQARKEKENAKRLNKL	235		
XX	XX	GPBP-interacting protein GIP130b.								
DE	DE									
XX	XX									
KW	KW	Goodpasture antigen binding protein interacting protein; GPBP; GIP130b;								
KW	KW	human; transcription factor; autoimmune disease; cancer;								
XX	XX	immunosuppressive; cytostatic.								
OS	OS	Homo sapiens.								
XX	PN	WO2003048193	A2.	416	HSKELRLVEKLO	KRMSELEKLEEA	FSRSKSECTQLHLNLEKEKNLT	TKDLLNELEVVKSR	475	
XX	PD	12-JUN-2003.								
XX	XX	05-DEC-2002; 2002WO-EP013802.								
PF	PF	07-DEC-2001;	2001US-0338287P.	476	VKELECSERLE	KAELSLKDDLT	KLKSFVTMLVDERKNMMEKIK	QOEERKV	DGLNKNFKVE	535
XX	XX	20-MAY-2002;	2002US-0382004P.	469	IKELEAIESR	LEKTEFTLKED	LTKLTITVNFVDERKTMSEK	LKKTED	KLQAASSQLQVE	528
PR	PR	(SAUS/) SAUS J.								
PA	PA	(REVE/) REVERT-ROS F.								
XX	XX	Saus J, Revert-Ros F;								
PI	PI									
XX	XX									
DR	DR	WPI; 2003-505281/47.								
XX	XX	N-PSDB; ACC83929.								
PT	PT	New Goodpasture antigen binding protein-interacting 90 and 130 kDa								
PT	PT	polypeptides, useful for diagnosing and/or treating disorders associated								
PT	PT	with the GIP90/130 polypeptide, such as autoimmune disorders and/or								
PT	PT	cancer.								
XX	XX	Claim 8; Page 78-83; 115pp; English.								
PS	PS									
XX	XX									
CC	CC	The present sequence is the protein sequence of GIP130b, a novel 130 kDa								
CC	CC	Goodpasture antigen binding protein (GPBP) interacting protein. The								
CC	CC	DOC/GIP90/130 mRNA family results from a complex diversification								
CC	CC	mechanism operating on the expression of the GIP90 gene on chromosome 3								
CC	CC	(3q12). GIP90/130 polypeptides of the invention interact with GPBP and								
CC	CC	are capable of aggregation. They can be used to modify GPBP-GIP90/130								
CC	CC	interactions; to modify GP90/130 aggregation, and to modulate gene								
CC	CC	expression. The invention provides GIP90/130 polypeptides, portions of								
CC	CC	them, antibodies, nucleic acid sequences, expression vectors and host								
CC	CC	cells, as well as methods for detecting GIP90/130 polypeptides and								
CC	CC	nucleic acids, and methods for treating an autoimmune disease or cancer								
CC	CC	by modifying the expression or activity of one or more GIP90/130								
CC	CC	polypeptides								
XX	XX	Sequence 1133 AA;								
SQ	SQ									
Query Match										
Best Local Similarity 40.3%; Score 2452.5; DB 6; Length 1133;										
Matches 542; Conservative 214; Mismatches 337; Indels 81; Gaps 19;										
QY	1	MRSRNOGGESSNGH	VSCPSSIISSDGGKGPSEDAKKNKANRKEE-DVMASGTIKRHLK	59	1049	VRKYN	SNANITTTEDNKI	HLGSGQFKRSPGPAAGVSPVITVRPVNVTAKEVSTGTVL	1108	
Db	1	MRSR--GSDTEGSAQK	KFRHT-----KGHSFGQPKMKHRQDKDSPESDV---IL	48	1045	QRNSN	SSSVTTEDNKI	HLGSPYMQA---VASPVRPASPSAPLQDNRTQGLINGALN	1101	
QY	60	PSGESEK----	KTKKSVELSKEDLIQLLSIMEGELQAREDVHMLRTEKTKPEVLEAHYG	115	1109	RSPRN	HLSSRPGASKV	TSTITITPVTTSTRGTQ	1142	
Db	49	PCPKAEKPHSGNGHQAEDLSRDDLFLLSILEGELQARDEVIGILKA	EKMDDLALAEAYG	108	1102	KT-----	TNKVTSSITITPTATPLPRQSQ	1125		


```
QY 368 GECGNSSLMAEVESLRKRVLEMEGKDEEITKTAEQRELKKKLQEEHHHSKELRLEVEKL 427
Db 121 GEYGNAGIMAEV-----EELIKMEEQCRDLNKNRLERETLQSKDFKLEVEKL 166
QY 428 QKRMSELEKLEEAFFSRKSECTQLHLNLEKEKNLT KDLLNELEVVKSRVKELECSERLE 487
Db 167 SKRIMALEKLEDAFNKSKQECYSLKCNLEKERMWTTKQLSQELESJKVRIKELEAIESRLE 226
QY 488 KAEKSLKDDLTCLKSFTVMLVDERKNMMEKIKQEERKVDGLNKNFKVEQGKVMDEVTEKLI 547
Db 227 KTEFTLKEDLTCLKLTVMFVDERKTMSEKLNKKTEDKLQAASSQLQVEQNKVTTVTEKLI 286
QY 548 EESKLLKXKSEMEKEYSLTKERDELMGKLRSEERSCELSCSVDLLKKRLDGIIEVER 607
Db 287 EETKRALKSKTDVEEKMYSVTKERDDDLKNKLKAEEEKGNDLLSRVNMMLKNRLQSLAEIAEK 346
QY 608 EINRGR----SCKGSEFTCPEDNKRILTLEIERLKKRLQQLQLEVVGEGLMKTEDYDQLE 663
Db 347 DFLKNKLNQDSGKSTTALHQENNKIKELSQEVEERLKLKDKMKAIEDDLMKTEDEYETLE 406
QY 664 QKFRTEQDKANFLSQOLEEIKHQMAKHAKIAEKEAVSQEAEHRFRLEEAKSRLDQAEV 723
Db 407 RRYANERDKAQFLSKELHVKMELAKYKLAEKTE-TSHEQWLFKRLQEEEAKSCHLSREV 465
QY 724 QALKEKIHLMNKEDQLSQVDYSVLQORFMEEETKKNMGREVLNLTKELELSKRYSR 783
Db 466 DALKEKIHAYMATEDLICHLQGDHVSCKKLNQOENNRDLGREIENLTKELEYRHFESK 525
QY 784 ALRPSGNRRMVDVPVASTGVQTEAVCGDAAEEETP-----AVFIRKSFQER-ENH 832
Db 526 SLRPSLNGRRISDPQVFSKEVQTEAV-----DNEPPDYKSLIPLERAVINGQLYESENQ 580
QY 833 IMSNLRQVGLKKPMERSVL-----DRYPPAANELTMRKSWIPWMRKRE----NGPSTPQ 883
Db 581 D-----EDPNDEGSVLSFKCSQSTPCPVN-----RKLWIPWMSKEGHLQNG----- 622
QY 884 EKGPRPNQGAGHPGELVLPKQGPPLHVRTPDHENSTATLEITSPTSE--EFFSSTTVI 941
Db 623 KMQTKPNANFVQPGDLVLSHTPGQPLHIKVTPDHVQNTATLEITSPTTESPHSYSTAVI 682
QY 942 PTLGNQKPRITIIPSPNVMSQPK-SADPTLGPERAMSPVTITISREKSPGEGRSAPAD 1000
Db 683 PNCGTPKQKITILQNASITPVKSKTSTEDLMNLEQGMSPITMATFARAQTPESCGSLTPE 742
QY 1001 RPAS.1004
Db 743 RTMS 746
RESULT 14
ID ABR92070
XX
AC ABR92070;
XX
DT 10-SEP-2003 (first entry)
XX
DE Human cervical cancer cell marker protein SEQ ID NO:48.
XX
KW Human; cervical cancer; cervical cancer marker; cancer therapy;
KW detection; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
PN WO2002101075-A2.
XX
PD 19-DEC-2002.
XX
PF 12-JUN-2002; 2002WO-US018638.
XX
PR 13-JUN-2001; 2001US-0298155P.
PR 13-JUN-2001; 2001US-0298159P.
```

```
PR 14-NOV-2001; 2001US-0335936P.
XX (MILL-) MILLENNIUM PHARM INC.
PA
XX
PI Schlegel R, Chen Y, Zhao X, Monahan JE, Kamatkar S;
PI Gannavarapu M, Glatt K, Hoersch S;
XX
DR WPI; 2003-156967/15.
DR N-PSDB; ACF12851.
XX
PT New isolated nucleic acid molecule useful for detecting, characterizing,
PT preventing and treating human cervical cancers, in various prognostic and
PT diagnostic assays, in pharmacogenomics and in monitoring clinical trials.
XX
PS Claim 4; Page 182-183; 386pp; English.
XX
CC ACF12828 to ACF12947 encode the human cervical cancer marker proteins (I)
CC given in ABR92047 to ABR92164. A higher level of expression of (I) than
CC normal indicates the presence of cervical cancer. Also described: (1) a
CC vector (II) containing (I); (2) a host cell (III) containing (I); and (3)
CC assessing (M1) whether a patient is afflicted with cervical cancer,
CC comprising comparing the level of expression of a marker in a patient's
CC sample, and the normal level of expression of the marker in a control non
CC -cervical cancer sample, where a significant increase in the level of
CC expression of the marker in the patient's sample relative to that in the
CC control sample is an indication that the patient is afflicted with
CC cervical cancer. (I) has cytostatic activity, and can be used in gene
CC therapy and in vaccines. (I) is useful in detecting, characterising,
CC preventing and treating human cervical cancers. (I) may also be used in
CC various prognostic and diagnostic assays, pharmacogenomics and in
CC monitoring clinical trials
XX
SQ Sequence 752 AA;
Query Match 27.1%; Score 1645.5; DB 6; Length 752;
Best Local Similarity 46.3%; Pred. No. 3.2e-76;
Matches 363; Conservative 140; Mismatches 216; Indels 65; Gaps 13;
QY 248 MLVDERQMHIEQLGQSQKVQDQLTQKLREEEEKLKAVTYKSKEDRQKLLKLEVDFFHKAS 307
Db 1 MVVDEQQRLTAQLTLQRQKIQLTNAKETHTTKLALAEARVQEEEQKATRLKELQTT 60
QY 308 RFSQHEEEMNAKLQESHNRQLRLKLVGLSORIEELEETNKSLOKAEELQELREKIAK 367
Db 61 KFHQDQDTIMAKLTNEDSQNRQLQKLAALSQIDLEETNRSRLKAEELQDIKEKISK 120
QY 368 GECGNSSLMAEVESLRKRVLEMEGKDEEITKTAEQRELKKKLQEEHHHSKELRLEVEKL 427
Db 121 GEYGNAGIMAEV-----EELIKMEEQCRDLNKNRLERETLQSKDFKLEVEKL 166
QY 428 QKRMSELEKLEEAFFSRKSECTQLHLNLEKEKNLT KDLLNELEVVKSRVKELECSERLE 487
Db 167 SKRIMALEKLEDAFNKSKQECYSLKCNLEKERMWTTKQLSQELESJKVRIKELEAIESRLE 226
QY 488 KAEKSLKDDLTCLKSFTVMLVDERKNMMEKIKQEERKVDGLNKNFKVEQGKVMDEVTEKLI 547
Db 227 KTEFTLKEDLTCLKLTVMFVDERKTMSEKLNKKTEDKLQAASSQLQVEQNKVTTVTEKLI 286
QY 548 EESKLLKXKSEMEKEYSLTKERDELMGKLRSEERSCELSCSVDLLKKRLDGIIEVER 607
Db 287 EETKRALKSKTDVEEKMYSVTKERDDDLKNKLKAEEEKGNDLLSRVNMMLKNRLQSLAEIAEK 346
QY 608 EINRGR----SCKGSEFTCPEDNKRILTLEIERLKKRLQQLQLEVVGEGLMKTEDYDQLE 663
Db 347 DFLKNKLNQDSGKSTTALHQENNKIKELSQEVEERLKLKDKMKAIEDDLMKTEDEYETLE 406
QY 664 QKFRTEQDKANFLSQOLEEIKHQMAKHAKIAEKEAVSQEAEHRFRLEEAKSRLDQAEV 723
Db 407 RRYANERDKAQFLSKELHVKMELAKYKLAEKTE-TSHEQWLFKRLQEEEAKSCHLSREV 465
QY 724 QALKEKIHLMNKEDQLSQVDYSVLQORFMEEETKKNMGREVLNLTKELELSKRYSR 783
Db 466 DALKEKIHAYMATEDLICHLQGDHVSCKKLNQOENNRDLGREIENLTKELEYRHFESK 525
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
OM protein - protein search, using sw model
Run on: September 7, 2004, 13:10:44 ; Search time 35 Seconds
(without alignments)
1787.734 Million cell updates/sec
Title: US-10-788-793-2
Perfect score: 6082
Sequence: 1 MRSRNQGGESSNGHVSCP.....KIELKKSAAASSTASLGGGKG 1212
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469	7.7	2662	4 US-09-595-684B-31	Sequence 31, Appl
2	449	7.4	1972	4 US-08-875-435B-3	Sequence 3, Appli
3	445	7.3	1972	4 US-08-875-435B-4	Sequence 4, Appli
4	436.5	7.2	2482	1 US-08-328-254-6	Sequence 6, Appli
5	433	7.1	3248	1 US-08-353-700-1	Sequence 1, Appli
6	433	7.1	3248	5 PCT-US95-16216-1	Sequence 1, Appli
7	429	7.1	1898	1 US-08-056-200-94	Sequence 94, Appl
8	429	7.1	1898	2 US-08-800-644-94	Sequence 94, Appl
9	420	6.9	3878	4 US-09-914-259-11	Sequence 11, Appl
10	413.5	6.8	1231	4 US-08-714-741-41	Sequence 41, Appl
11	409.5	6.7	1939	4 US-09-310-187A-1	Sequence 1, Appli
12	407.5	6.7	1886	4 US-08-938-105-3	Sequence 3, Appli
13	407.5	6.7	2101	1 US-08-466-390-4	Sequence 4, Appli
14	407.5	6.7	2101	1 US-08-470-950-4	Sequence 4, Appli
15	407.5	6.7	2101	1 US-08-467-781-4	Sequence 4, Appli
16	407.5	6.7	2101	1 US-08-195-487-4	Sequence 4, Appli
17	407.5	6.7	2101	2 US-08-483-924-4	Sequence 4, Appli
18	407.5	6.7	2101	3 US-09-452-294-1	Sequence 1, Appli
19	407.5	6.7	2101	5 PCT-US93-06160-4	Sequence 4, Appli
20	401.5	6.6	2954	4 US-09-150-867-1	Sequence 1, Appli
21	395	6.5	1388	4 US-09-572-191-2	Sequence 2, Appli
22	395	6.5	1388	4 US-09-723-262-2	Sequence 2, Appli
23	395	6.5	1388	4 US-09-723-219-2	Sequence 2, Appli
24	393	6.5	1695	4 US-09-866-108A-15753	Sequence 15753, A
25	389	6.4	2125	4 US-09-919-172-29	Sequence 29, Appl
26	380	6.2	976	3 US-09-104-324B-4	Sequence 4, Appli
27	379	6.2	1312	2 US-08-687-080-51	Sequence 51, Appl

28	378	6.2	1312	2 US-08-592-126-148	Sequence 148, App
29	378	6.2	1312	4 US-09-168-595-148	Sequence 148, App
30	376.5	6.2	1581	4 US-09-866-108A-15754	Sequence 15754, A
31	375.5	6.2	1354	3 US-08-685-871-2	Sequence 2, Appli
32	374.5	6.2	2568	4 US-09-866-108A-3	Sequence 3, Appli
33	369.5	6.1	1388	4 US-09-976-594-296	Sequence 296, App
34	367.5	6.0	1388	2 US-08-685-576-4	Sequence 4, Appli
35	366.5	6.0	1388	2 US-08-685-576-1	Sequence 1, Appli
36	359	5.9	973	4 US-09-392-714-24	Sequence 24, Appl
37	356	5.9	885	2 US-08-533-306A-4	Sequence 4, Appli
38	356	5.9	885	2 US-08-742-923A-4	Sequence 4, Appli
39	354.5	5.8	2468	4 US-09-976-594-726	Sequence 726, App
40	350.5	5.8	816	2 US-08-533-306A-6	Sequence 6, Appli
41	350.5	5.8	816	2 US-08-742-923A-6	Sequence 6, Appli
42	350.5	5.8	1093	5 PCT-US93-03077-1	Sequence 1, Appli
43	342.5	5.6	1530	4 US-09-976-594-736	Sequence 736, App
44	341	5.6	977	4 US-09-010-147B-18	Sequence 18, Appl
45	338	5.6	8991	4 US-08-714-741-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; TITLE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-31

Query Match	7.7%;	Score 469;	DB 4;	Length 2662;
Best Local Similarity	20.3%;	Pred. No. 2.2e-18;		
Matches	279;	Conservative 267;	Mismatches 455;	Indels 376; Gaps 62;
QY	33	SEDAKKNKANRKEED-----VMASTIKRHLKPSGESEKTK-----KSVE	73	
Db	1381	SQSQEQSLNMKEKDNETTKIVSEMEQFKPKDSALLRIEIMGLSKRLQESHDMKSA	1440	
QY	74	LSKEDLIQLLSIMEGEL-QAREDVIHMLRTE-KTKPEVLEAHYGSAPKVLRLVLRDAI	131	
Db	1441	KEKDDLQRLQEVLSQSDQLKENIKEIVAKHLETEELKVAHCCLKEQETINELR---V	1497	
QY	132	LAQEKSIGEDVVEKPISEL-DRLEEKQKETVYRMLEQLLLAE-----KCHR	176	
Db	1498	NLSEKETEISTIQLEAINDKLQNIQEIYKEB-EQLNIKQISEVQENVNKLQFKEHR	1556	
QY	177	R---TVYELENEKHKHTDYMNKSDDFTNLLQERERLKKLLEQEKAYQARKE--KENAK	230	
Db	1557	KAKDSALQSIESKMLELTNRLQESQEEIQIMIKEEMKRV---QEQALQIERDQLKENTK	1613	
QY	231	RL-NKLRLDELVK---LKSFALMLVDERQMHIQL--GLQSQKVQ-----DLTQKL	274	
Db	1614	EIVAKMKESQEKYQFLKMTAVNETQKMCIEIHLKEQFETQKLNLENIETENIRLTQIL	1673	

;;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-328-254-6

Query Match 7.2%; Score 436.5; DB 1; Length 2482;
Best Local Similarity 20.0%; Pred. No. 1.4e-16;
Matches 275; Conservative 235; Mismatches 471; Indels 395; Gaps 54;

QY 69 KKSVELSK-----EDLIQLLSIMEG-----ELQAREDV----- 96
Db 1205 KENS DLSEKLEYFSCDHQELLQRVETSEGLNSDLEMHADKSSREDIGDNVAKVNDSWKER 1264
QY 97 -----IHMLRTEKTKPE----VLEAHYGSAAPEKVLRLVLRD-----AILAQEKS 137
Db 1265 FLDVENE LSRIRSEKASIEHEALYLEADLEV VQTEKL--CLEKDNENKQKVIVCL EEBLS 1322
QY 138 I-----GE-DVYEKPISELDRLEBKQKETYRRMLEQLLLAEKCHRRTVYELENKX 187
Db 1323 VVTSERNQLRGELDTMSKKT TALDQLSEKMK EKTQEL--ESHQSECLHCIOVAEAE-VKE 1379
QY 188 KHTDYMNKSD DFTNLLEQERERLKKLLEQEKAYQA----RKEKEN-AKRLNKLRLDELVK- 241
Db 1380 KTELLQTLSSDVS ELLKDKTKHLQEKLSLEKDSQALS LTKCELENQIAQLNKEKELLVKE 1439
QY 242 -----LKS FALMLVDERQMHI EQGLQSQKVQDLTQKLREEE EKLKA 283
Db 1440 SESLQARLSESDYEKLVN VSKALEAALVEK-----GEFALRLSSTQEEVHQLRRGIEKL RV 1494
QY 284 VTYKSKEDRQKLLKLEVD FEHKASRF SQEHEEMNAKLANOESHNRQIRLKLVLGLSQRIEE 343
Db 1495 ---RIADEKKQLHI-----AEKLKERERENDSLKDKVENLE--RELQMSEENQELVILD 1544
QY 344 LEETNKS LQKAAEEELQBLREKIAKGECGNSSSLMAEVESLRKRVLEMEGKDEEITTKTEAQC 403
Db 1545 AENSKAEVETLKTQIEEMARSLKVFELDLVTLRSEKENLTKQIQEKQGLSELDKLLSSF 1604
QY 404 RE-LKKKLOEEHHSEKLRLEVEKLOKRMSELEKL----- 439
Db 1605 KSLLEKEQAEIQIKEESKTA VEMLQNQLKELNEAVALCGDQEI MKATEQSLDPPIEEE 1664
QY 440 AFSRSKSECTQLHLNLEKEKNL-----TKDLLNELEVVKSRV----KELECSESRLKAE 490
Db 1665 HQLRNSIEKLRARLEADEKKQLCVLQOLKESEHHDLLKGRVENLERELEIARTNQEHAA 1724

QY 491 L-----SLKDDLTKLKSFTVMLVDERKNMMEKIKQEERKVDGL---NK 530
Db 1725 LEAENSKGEVETLKAIEGMTQSLRGLELDVVTIRSEKENLTNQLQKEQERISELEIINS 1784
QY 531 NFKVEQGVMDVTEKLJIEESKLLKLKSEMEEKEYSLT-----KERDELMGKLRSE 581
Db 1785 SF-----ENILQE-----KEQEKVQMKESKSTAMEMLOTQLKELNERVAALHND 1828
QY 582 EERSCE----LSCSVDLLK---KRLDGI EEEV-----REINRGRSCKGS 618
Db 1829 QE-ACKAKEQNLS SQVECLELEKAQLLQGLDEAKNNYIVLQSSVNGLIQEVEDGK----- 1882
QY 619 EFTCPEDNKIRELTLEIERLKKRLQQLLEVVEGDLMKTEDEYD-----OLEQ 664
Db 1883 -----QKLEKKDEEISR LKNQIQOEQLVSKLSQVEGEHQWLWKEQNLELRLNLTVELEQ 1935
QY 665 KFRTEQDKANFLSQOLEEIKH-----OMAKHKAIEK-GEAVSQEAE LRRHFR 711
Db 1936 KIQVLQSKNASLQDTLLEV LQSSYKNLENELELT TMDKMSFVEKVNMKTAKETELQREHME 1995
QY 712 EEA KSRDLQAEVQALKEK-----IHELMNKEDQLSQLDYDYSVLQQRF---MEEET 759
Db 1996 MAQKTAE LQEELSGEKNRLAGELQLLLEI KSSKDQLKELTLENSELKKS LDCMHKDQVE 2055
QY 760 KNKNMGREVLNLTKELELSKRYSRALRPSGNGRRMVDVPVASTGVQTEAVCGDAAEEETP 819
Db 2056 KEGKVREEIAEYQ LRLHEAEKKHQALLD TNKQYEVEIQTYREKLT SKEECLSSQKLE-- 2113
QY 820 AVFIRKSFQEE--NHI-----MSNLRQVG-LKKPMERS-SVLDRYPPAANE 861
Db 2114 -IDLLKSSKEELNNSL KATTQILEELKKTMDNLKYVNLKKNERAOGKMKLLIKSKCQ 2172
QY 862 LTMRKSWIPWMRKRENGPSTPQEKGPRPNQAGHPGELVLAPKQGQPLHIRV---TPDHE 918
Db 2173 LEEEK E--ILQKELSQLAAQEK-----QKTGTVM DTKVDEL TTEIK 2212
QY 919 NSTATLEITSPTSEEFFSS-----TTVIPTLGNQKPRITIIIPSPNVM 960
Db 2213 ELKETLEEKTKEADEYLDKYCSLLISHEKLEKAKEMLE TQVAHLCSQ----- 2260
QY 961 SQPKSADPTLGP ERAMSPVTITISREKSPEGGRSAFADR-----PASPI 1006
Db 2261 SKQDSRGSP L LGPV-VPGPSIPSVTEKRLSSGONKASGRQRSSGIWENG GGPPTPATPE 2319
QY 1007 QIMTVSTSAAPTEIAVSPESQ-----EVP MGR-----ILKVTPEKQT 1044
Db 2320 SPSKSKKAVMSGIHPAEDTEGTEFEPEGLPEVVKKG FADIPTGKTSPIYILRRITMATRT 2379
QY 1045 VPAPVRKYNSNANI TTEDNKIHLG SQFKRSPGPAAGVSPVITVRPVNVTAKEKEVST 1104
Db 2380 SPRLAAQKLALSP LSLGKEN-----LAESSKPTAGGSR S-----QKVKAQRSPVDS 2426
QY 1105 GTVLRSPR-----NHL-----SSRPGASKVTSTITITPVTTSSTRGTQSVSGQ 1147
Db 2427 GTILREPTTKSVPVN NLPERSPTDSPREGLRVKRGRVLVPSPKAGLESKGS ENCKVQ 2482

RESULT 5
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA

STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,700
FILING DATE: 09-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: REED, JANET E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: HUMAN
US-08-353-700-1

Query Match 7.1%; Score 433; DB 1; Length 3248;
Best Local Similarity 20.1%; Pred. No. 3.1e-16;
Matches 270; Conservative 228; Mismatches 454; Indels 388; Gaps 53;
QY 69 KKSVELSK-----EDLIQLLSIMEG-----ELQAREDV----- 96
DB 1933 KENS DLSEKLEYFSCDHQELLQRVETSEGLNSDLEMHADKSSREDIGDNVAKVNDWSKER 1992
QY 97 -----IHMLRTEKTPE-----VLEAHYGSARPEKVLRLVLRD-----AILAQEKS 137
DB 1993 FLDVENELSRIRSEKASIEHEALYLEADLEVTVQTEKL--CLEKDNENKQKVICLEEEELS 2050
QY 138 I-----GE-DVYEKPISELDRLEEKQKETYRRMLEQLLLAEEKCHRRTVVELENEKH 187
DB 2051 VVTSERNQLRGELDTMSKKTALDQLSEKMKETQEL--ESHQSECLHCICQVAEAE-VKE 2107
QY 188 KHTDYMNKSDFTNLLEQERERLKKLLEQEKAYQA----RKEKEN-AKRLNKLRLDELVK- 241
DB 2108 KTELLQTLSSDVSELLKDKTHLQELKQSLKESQALSQALSLTKCELENQIAQLNKEKELLVKE 2167
QY 242 -----LKSFALMLVDERQMHIQGLQSQKQVQDLTQKLREEEEEKLKA 283
DB 2168 SESLQARLSESDYEKLVNSKALEALVEK-----GEFALRLSSTQEEVHQLRRGIEKLRV 2222
QY 284 VTYKSKEDRQKLLKLEVDFFEHKASRFSQEHHEEMNAKLANQESHNRQLRLKLVGLSQRIEE 343
DB 2223 ---RIEADKKQLHI----AEKLKERERENDSLKDKVENLE---RELQMSEENQELVILD 2272
QY 344 LEETNKSLOKAEELQELREKIAKGECGNSSLMAEVESLRKRVLEMEGKDEEITKTEAQ 403
DB 2273 AENSKAEVETLKTQIEEMARSLKIFELDLVTLRSEKENLTKQIQEKQGGQSLSELDKLLSSF 2332
QY 404 RE-LKKKLQEEHHSKELRLEVEKLQKRMSELEKL-----EE 439
DB 2333 KSLLEEKEQAEIQIKESKTAVEMLQNLKELNEAVALCGDQEIIMKATEQS LDPPIEEE 2392
QY 440 AFSRSKSECTQLHLNLEKEKNL-----TKDLLNELEVWVKS RV---KELECSSESRLKAE 490
DB 2393 HQLRNSIEKRLARLEADEKKQLCVLQQLKESEHADLLKGRVENLERELEIARTNQEHAA 2452
QY 491 L-----SLKDDLTCLKSFTVMLVDERKNMMEKIKQEERKVDGL---NK 530

DB 2453 LEAENSKGEVETLKAKIEGTMQSLRGLDLDVVVIRSEKENLTNELQKEQERISELEILNS 2512
QY 531 NFKVEQGVMDVTEKLIIEESKLLKLLKSEMEEEKYSLT-----KERDELMGKLRSE 581
DB 2513 SF-----ENILQE-----KEQEKVQMKESSTAMEMLQTLQKELNERVAALHND 2556
QY 582 EERSC-----LSCSVDLLK---KRLDGIEEBE-----REINRGRSCKGS 618
DB 2557 QE-ACKAKEQNLSSQVEGLEKAQLLQGLDEAKNNYIVLQSSVKGLIQEVEDGK----- 2610
QY 619 EFTCPEDNKIRELTLEIERLKKRLQQLLEVVEGDLMKTEDEYD-----QLEQ 664
DB 2611 -----QKLEKKDEEISRLKNQIQDQEQVLVSKLSQVEGEHQLWKEQNLRLNLTVELEQ 2663
QY 665 KFRTEQDKANFLSQOLEEIKH-----QMAKHKAIEK-GEAVSQBAELRHRFRL 711
DB 2664 KIQVLQSKNASLQDTLEVLOSSYKNLENELELTCKMDKMSFVEKVNKMTAKETELQREMHE 2723
QY 712 EEAKSRLQAEVQALKEK-----IHELMNKEDQLSOLQVDYSVLQQRFF-----MEEET 759
DB 2724 MAQKTAELQEELSGEKNRLAGELQLLLLLEEIKSSKDQLKELTLENSELKSLDCMHKQDVE 2783
QY 760 KNKNMGREVLNLTKELELSKRYSRALRPSGNRRMRVDVPVASTGVQTEAVCGDAAEETP 819
DB 2784 KEGKVREEIAEYQRLHEAEKKHQALLDNTNKQYEVEIQTYREKLTSCKEELSSQKLE-- 2841
QY 820 AVFIRKSFQEE--NHI-----MSNLRQVG-LKKPMERS-SVLDRYPPAANE 861
DB 2842 -IDLLKSSKEELNNSLKATTQILEELKTKMDNLKYNVQLKKENERAQGMKLLIKSCKQ 2900
QY 862 LTMKRSWIPWMRKRENGPSTPOEKGPRPNQAGHPGELVLAPKQGQPLHVR--TPDHE 918
DB 2901 LEEKE--ILQKELSOLQAAQEK-----QKTGTVMDTKVDELTTTEIK 2940
QY 919 NSTATLEITSPTSEEFSS-----TTVIPTLGNQKPRITILIPSPNVM 960
DB 2941 ELKETLEEKTKADEYLDKYCSLLISHEKLEKAKEMLETOVAHLCSQ----- 2988
QY 961 SQPKSADPTLGPERRAMSPVTITTSIREKSPEGGRSAFADR-----PASPI 1006
DB 2989 SKQDSRGSPLLPV-VPGSPIPSPVTEKRLSSGQNKASGKRQRSSGIWENGRGPTPATPE 3047
QY 1007 QIMTVSTSAAPTEIAVSPESQ-----EVPMGRT---ILKVTPEKQT 1044
DB 3048 SFSKSKKAVMSGIHPAEDTEGTEFEPEGLPEVVKGFADIPTKTSPYILRRTTMA TRT 3107
QY 1045 VPAPVRKYNSNANIITTEDNKIHLGSOFKRSPGPAAGVSPVITVRPVNVTAKEVST 1104
DB 3108 SPRLAAQKLALSPLSLGKEN-----LAESSKPTAGGSR-----QKVKVAQRS PVDS 3154
QY 1105 GTVLRSPR-----NHLSSR 1118
DB 3155 GTILREPTTKSVPVNNLPER 3174

RESULT 6
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-1

Query Match 7.1%; Score 433; DB 5; Length 3248;
Best Local Similarity 20.1%; Pred. No. 3.1e-16;
Matches 270; Conservative 228; Mismatches 454; Indels 388; Gaps 53;

Qy 69 KKSVELSK-----EDLIQLLSIMEG-----ELQAREDV----- 96
Db 1933 KENSLSSEKLEYFSCDHQELLQORVETSEGLNSDLEMHADKSSREDIGDNVAKVNDWSEKER 1992

Qy 97 -----IHMLRTEKTKPE-----VLEAHYGSAPPEKVLRLVLRD-----AILAQEKS 137
Db 1993 FLDVENELSRIRSEKASIEHEALYLEADLEVVVQTEKL--CLEKDNENKQKVIVCLLEEELS 2050

Qy 138 I-----GE-DVYEKPISELDRLEEKQKETYRRMLQLLLAEKCHRRTVYELENEKH 187
Db 2051 VVTSERNQLRGELDTMSKKTALDQLSEKMKKEKTQEL--ESHQSECLHCIOVAEAE-VKE 2107

Qy 188 KHTDYMKNKSDFTNLLQEORERLKLLEQEKAYQA-----RKEKEN-AKRLNKLRLDELVK- 241
Db 2108 KTELLQTLSSDVSELLKDKTHLOEKLSLEKDSQALSCLKELQENQIAQLNKEKELLVKE 2167

Qy 242 -----LKSFALMLVDERQMHIEQLGLQSQKVQDLTQKLREEEEKLLKA 283
Db 2168 SESLQARLSSESDYEKLNVSKALEALVEK-----GEFALRLSSTQEEVHQLRGRIEKLRV 2222

Qy 284 VTYKSKEDROKLLKLEVDPEHFKASRFSQEEHEEMNAKLANQESHNRQLRLKLVGLSQRIEE 343
Db 2223 ---RIEADKKQLHI-----AEKLERERENDSLKQKVLENLE---RELQMEENQELVILD 2272

Qy 344 LEETNKSLOKAEELQELREKIAKGECGNSSILMAEVESLRKRVLEMGKDEEITKTEAQC 403
Db 2273 AENSKAEVETLKTQIEEMARSLKIFELDLVTLRSEKENLTKQIQEKQQLSELDKLLSSF 2332

Qy 404 RE-LKKKLQEEHHSKELRLEVEKLQKRMSELEKL-----EE 439
Db 2333 KSLLEKEQAEIQIKESKTAVEMLQNLKELNEAVALCGDQEIIMKATEQSLDPPPIEEE 2392

Qy 440 AFSRSKSECTQLHLNLEKEKNL-----TKDLLNELEVVKSRV-----KELECSERLEKAE 490
Db 2393 HQLRNSIEKLRLARLEADEKKQLCVLQQLKESEHHDLLKGRVENLERELEIARTNOEHAA 2452

Qy 491 L-----SLKDDLTCLKSFTVMLVDERKNMMEKIKQEEKVDGL---NK 530
Db 2453 LEAENSKGEVETLKAKIEGMTQSLRGLELDVVITRSEKENLTNELQKEQERISELEIINS 2512

Qy 531 NFKVEQGVMDVTEKLIIEESKLLKLSKEMEKEYSLT-----KERDELMGKLRS 581
```

```

Db 2513 SF-----ENILQE-----KEQEKVQMKEKSSSTAMEMLTQQLKELNERVAALHND 2556
Qy 582 EERSCE-----LSCSVDLLK-----KRLDGIEEVE-----REINRGRSCKGS 618
Db 2557 QE-ACKAKEQNLSQVECLEKAQLLQGLDEAKNNYIVLQSSVKGLIQEVEDGK----- 2610
Qy 619 EFTCPEDNKIRELTLEIERLKKRLQQLLEVVEGDLMKTEDEYD-----QLEQ 664
Db 2611 -----QKLEKKDEEISRLKNQIQDQEQVLVSKLSQVEGEHQLWKEQNLELRLNLTVELEQ 2663
Qy 665 KFRTEQDKANFLSQLEBIKH-----QMAKHAIEK-GEAVSQEAEALRHRFRL 711
Db 2664 KIQVLQSKNASLQDTLLEVLQSSYKYNLENELELTMDKMSFVEKVNKMTAKETELQREMHE 2723
Qy 712 EEAQRDLQAEVQALKEK-----IHELMNKEDQLSQLQVDYSVLQQRF---MEEET 759
Db 2724 MAQKTAEQLQEELSGEKNRLAGELQLLLEEIKSSKDQLKELTLENSELKSKSLDCMHKQDVE 2783
Qy 760 KNKNMGREVLNLTKELELSKRYSRALRPSGNGRRMVDVPVASTGVQTEAVCGDAAEETP 819
Db 2784 KEGVREEIAEYQRLHEAEKHHQALLDNTKNQYEVETQTYREKLTSKBECLSSQKLE-- 2841
Qy 820 AVFIRKSFQEE--NHI-----MSNLRQVG-LKKPMERS-SVLDRYPPAANE 861
Db 2842 -IDLKSSKEELNNSLKATTQILEELKTKMDNLKYVNLKKNENAKQKMKLLIKSKCKQ 2900
Qy 862 LTMKRSWIPWMRKRENGPSTPQEKGRPNQAGHPGELVLAPKQGPPLHIRV---TPDHE 918
Db 2901 LEEKEB---ILQKELSQLQAAQEK-----QKTGTVMDDTKVDELTTTEIK 2940
Qy 919 NSTATLEITSPTSEEFSS-----TTVIPTLGNQKPRITIIPSPNVN 960
Db 2941 ELKETLEEKTKEADEYLDKYCSLLISHEKLEKAKEMLETOVAHLCSQQ----- 2988
Qy 961 SQPKSADPTLGPERSAMSPVTITITISREKSPESGGRSAFADR-----PASPI 1006
Db 2989 SKQDSRGSPLLPV-VPGSPIPSVTEKRLSSGQNKASGRQRSSGIWENGRGPTPATPE 3047
Qy 1007 QIMTVSTSAAPTEIAVSPESQ-----EVPMGRT-----ILKVTPEKQT 1044
Db 3048 SFSKSKKAVMSGIHPAEDTEGTEPEPEGLPEVVVKGFADIPTGKTSPIYLLRRTTMATRT 3107
Qy 1045 VPAPVRKYNSNANIITTEDNKIHLGSGQFKRSPGPAAGVSPVITVRPVNVTAEKEVST 1104
Db 3108 SPRLAAQKLALSPLSLGKEN-----LAESSKPTAGGSR-----QKVKAQRSPVDS 3154
Qy 1105 GTVLRSPR-----NHLSSR 1118
Db 3155 GTILREPTTKSVPVNVLPER 3174

RESULT 7
US-08-056-200-94
; Sequence 94, Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/056,200
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-056-200-94

Query Match          7.1%; Score 429; DB 1; Length 1898;
Best Local Similarity 23.2%; Pred. No. 2.7e-16;
Matches 216; Conservative 168; Mismatches 376; Indels 172; Gaps 35;

QY 34 EDAKNKANRKEEDVMASGTIKRHLKPSGESEKTKKSVLSKEDLIQLLSIMEGELQAR 93
Db 344 ERREQQLRREQEERREQQLRREQEERREQQLRREQEERREQQLRREQQLR 403
QY 94 EDVIHMLRTEKTKPEVLEAHYGSAPKVLK---VLHRDAILAQEKSIGEDVYEKPISEL 150
Db 404 RE--QQLRREQ-----QLRREQQLRREQQLRREQQLRREQQLRREQE 450
QY 151 DRLEEKQKETYRRMLEQLLAEKCHRRRTVYELENEKHKHTDYMNKSDDFTNLLEQERERL 210
Db 451 ERHEQKHEQERR--EQLKREQEERRDVLKREEETERHEQERRKQQLKRDQEEERRERW 507
QY 211 KKLLEQEKAYQARKEKENAKRLNKLRLDELVKLSFALMLVDERQMHIQGLSQKVQDL 270
Db 508 LKLEEEERREQERREQQLRREQEERRE-QRLK-----RQEEERL---QQLRSE 554
QY 271 TQKLREEEKLKAVTYKSKEDRQKLLKLEVDFEHKASRFSQEHEEMNAKLANQESHNRQL 330
Db 555 QQLRREQEERLEQLLKREBEK-----LEQERREQRLKREQEERRDQLLKREBEERRQ 607
QY 331 RLKL---VGLSQRIE-----ELEETNKSLOKAEELQELREKIAKGECGNSLMAEVE 380
Db 608 RLKREQEERLEQLKREVERLEQEERRDERLKRREEPEERRHELLKSE----- 656
QY 381 SLRKRVLEMEGKDEEITKTEAQCRELKKLQEEEEHHSKELRL-----EVEKLOKRMSELEK 436
Db 657 -----EQEERRHEQLRREQEERREQRLKREBEERLEQLKREHEEERREQELAE-EE 708
QY 437 LEEAFSRKSECTQLHLNLEKEKNL--TKOLLNELEVVKSRVKELECSSESRLKAELSLK 494
Db 709 QEQARERIKSRIPKWQWQLESEADARQSKVLLEAPQAGRAEAPQ-EQEEKRRRESELQWQ 767
QY 495 DDLTKLSFTVMLVDERKNMMEKIKQEEERKVD---GLNKNFKVEQGV-MDVTKEKLEES 550
Db 768 E-----EERARHQ--QEEERRDFTWQWQAEKSERGRQRLSARPLREQR 812
QY 551 KKLKLKXSEMEEKEYSLTKERD-ELMGKLRSEERSECSLSCVDLLKKR-----LD 600
Db 813 ERQLRAEERQREQRFLPEEEKEQGRQRORREREXELQFLEEEQLORRERARQQLQEEED 872
QY 601 GI-EEVEREINRGRSCKGSEFTCPEDNKIRELTLEI-----ERLKKRLQQLVEVVGDLMK 654
Db 873 GLQEDQERRRQEQRRDQKRWQLEEEERKRRRHTLYAKPALQEQRLKKEQQLQEEEEEELQ 932
QY 655 TEDE---YDQLEQKFRTE---QDKANFLSQLE-----EIKHQMAKHKAIEKGE--AVS 700
Db 933 EEREKRRRQEQERYHEEEQLQEEEEEQLLREEREKRRRQERERYRKDKKLQKKEEQLLG 992
```

```

QY 701 QEAELRHRFRLEEAKSRD---LQAEVQALKE-----KIHELMNKEDQLS 741
Db 993 EEPEKRRR-QEREKKYREEEELQEEEEQLLREEREKRRRQEWERYKKKDELQEEEQLL 1051
QY 742 QLQVDYSVLQQR---FMEEE---TKNKNMGRE-----VLNLTKELELSKRYSRAL 785
Db 1052 REEREKRRRLQERERYREEEELQEEEEQLLGEETRRRRQELERQYRKEEELQEEEQLL 1111
QY 786 RPSGNGRRMVDVPVASTGVQTEAVCGDAAEEETPAVFIRKSFOEENHIMSNLRQVGLKKP 845
Db 1112 REEPEKRRR-----QERERQCREEEELQEEEEEQLLREEREKRRRQELERQYREEEE 1162
QY 846 MERSSVLDRYPPAANELTMRKSWIPWMRKREN 877
Db 1163 LQRQKRQRYRDEQSDLKQWEP---EKEN 1191

RESULT 8
US-08-800-644-94
; Sequence 94, Application US/088000644
; Patent No. 5958752
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Methods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,644
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/056,200
; FILING DATE: 30-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-800-644-94
```

```

Query Match          7.1%; Score 429; DB 2; Length 1898;
Best Local Similarity 23.2%; Pred. No. 2.7e-16;
Matches 216; Conservative 168; Mismatches 376; Indels 172; Gaps 35;

QY 34 EDAKNKANRKEEDVMASGTIKRHLKPSGESEKTKKSVLSKEDLIQLLSIMEGELQAR 93
Db 344 ERREQQLRREQEERREQQLRREQEERREQQLRREQEERREQQLRREQQLR 403
```

```
QY 94 EDVIHMLRTEKTPVLEAHYGSABPEKVL R---VLHRDAIIAQEKSIGEDVVEKPISEL 150
Db 404 RE--QQLRREQ-----QLRREQQLRREQQLRREQQLRREQQLRREQQLRREQE 450
QY 151 DRLEBKQKETRYRMLEQALLAEKCHRRTVYELENEKHGHTDYMNKSDDFTNLLEQERERL 210
Db 451 ERHEQKHEQERR---EQRLKREQEERRDWLKRBEETERHEQERKQQLKRQDBEERRERW 507
QY 211 KLLLEQEKAYQARKEKENAKRLNKLDELVKLSFALMLVDERQMHIEQLGLGLOSQKVQDL 270
Db 508 LKLEEEERREQERREQQLRREQERRE-QRLK-----RQEEERL---QQLRSE 554
QY 271 TQKLREEEKLVKATYKSKEDRQKLKLEVDFFHKASRFSQEHHEMNAKLANQESHNRQL 330
Db 555 QQLRREQEERLEQLLKREEKR-----LEQERREQRLKREQEERRDQLLKREEERRQQ 607
QY 331 RLKL---VGLSORIE-----ELEETNKSLOKABEEELQELRBKIAKGECGNSSLMAEVE 380
Db 608 RLKREQEERLEQRLKREEVERLEQEEERRDERLKRPEEPERHELLKSE----- 656
QY 381 SLRKRVLEMEGKDEEITKTQAQRELKKKLQBEHHHSKELRL---EVEKLQKRMSELEK 436
Db 657 -----EQEERHLEQRLRREQERREQRLKREBEERLEQRLKREHEERREQELAE-EE 708
QY 437 LEEAFSRKSECTQHLNLEKKNL--TKDLLNELEVVKSRVKELECSERLEKAELS LK 494
Db 709 QEQARERIKSRIPKQWQWQLESEADARQSKVLEAPQAGRAEAPQ-EQEEKRRRESELOWQ 767
QY 495 DDLTKLSFTVMLVDERKNMMEKIKQEEERKVD---GLNKNFKVEQGKV-MDVTEKLIIES 550
Db 768 E-----EERAHROQ--QEEERQDRFTWQWQAEKSERGQRLSARPPLEQR 812
QY 551 KKLKLLKSEMEKEYSLTKERD-ELMGKLRSEERSCELSVDLLKKR-----LD 600
Db 813 ERQLRAEERQEQERFLPEEEKEQGRQREREREKELQFLBEEELQRRERRAQQLQEEED 872
QY 601 GI-EEVEREINRGRSCKSGSEFTCPEDNKIRLTLEI-----ERLKKRLQOLEVVEGDLMK 654
Db 873 GLQEDQERRRQEQRRDQKRWQLEEEERKRRRHTLYAKPALQEQQLRKEQQLLQEEEEELQ 932
QY 655 TEDE---YDQLEQKFRTE---QDKANFLSQOLE-----BIKHQMAKHAIEKGE--AVS 700
Db 933 EEREKRRRQEQERYREBELQOEELQOEELQOEELQOEELQOEELQOEELQOEELQ 992
QY 701 QEAELRHRFRLEAKSRD---LOAEVQALKE-----VLNLTKELELSKRYSRAL 785
Db 1052 REEREKRLQERERQYREBELQOEELQOEELQOEELQOEELQOEELQOEELQOEEL 1111
QY 786 RPSGNGRMRVDVPVASTGVQTEAVCGDAABEETPAVFIRKSFQEEHIMS NLRQVGLKXP 845
Db 1112 REEPEKRRR-----QERERQCREEEELQOEELQOEELQOEELQOEELQOEELQOEEL 1162
QY 846 MERSSVLDRYPPAANELTMRKSWIPWMMRKREN 877
Db 1163 LORQKRQORYDEDDQRSDLKWQWEP---EKEN 1191
```

```
RESULT 9
US-09-914-259-11
; Sequence 11, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
```

```
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-914-259-11

Query Match      6.9%; Score 420; DB 4; Length 3878;
Best Local Similarity 20.8%; Pred. No. 2.1e-15;
Matches 238; Conservative 185; Mismatches 327; Indels 394; Gaps 47;

QY 29 GKGPSSEDAKK-----NKANRKEEDVMASGTIKRHLK 59
Db 41 GQSPSKQKKRKTSSSKHDVSAHDLNIDQSQCNEMYINSSQVESTIPESTIMRTLH 100
QY 60 PSGESEKTKKSVLSKE----- 77
Db 101 SGEITSHEQGFVSVELESEISTTADDCSEVNGCSFVMRTGKPTNLLREBEFGVDDSYSEQ 160
QY 78 ---DLIQLLSIMEGELQAREDVIHMLRTEKTKPEVLEAHYGSAPK V---LRVLHRDA 130
Db 161 GAQDSPHLEMMSELAGKQHEIEELNREL--EEMRVTYGTGLQQLQEFEEAAIKQRDG 217
QY 131 ILAQEKSIGEDVYEKPISELDRLEEKQKETYRRMLE-----QLLLAEKCHRR 178
Db 218 IITQ-----LTANLQOARREKQDETMRFELELTSQSKLQIQFQQLQASSETLRNS 266
QY 179 VY-----ELENEKHKHTDYMNKSDDFTNLLEQERERLK----- 211
Db 267 THSSTAADLLQAKQIILTHQQOLEEQDHLLLEDYQKKKEDFTMQISFLQEKIKVYEME QDK 326
QY 212 -----KLLEQEKAYQARKEKENA--KRLNKLRLDELV---KLK 243
Db 327 KVENSKBEEIOKETIIEELNTKIIEBKKTLELKDCLTTADKLLGELQEQIVQKNQEI K 386
QY 244 SFALMLVDERQMH-----IEQLGLQSQK-----VQDLTQKLREEEEEKL 282
Db 387 NMKLELTSKQKERSQSEEI KQLMGTVVEELQKRNHKDSQFETDIVQRMEOETQ RKLQELR 446
QY 283 A-----VTYKSKEDRQKLLKLEVDFEHKA SRFQSQHEHEMNAKLANQESHNRQLRL 332
Db 447 AELDEMYGQQIVQMKQELIRQHMAQM---EEMKTRHKGEMENALRSYSNITVNEDQIKL 502
QY 333 KLVGLSQRIELEEETNKSLOKAEELQELREKIAKGECSGNSSLMAEVESLRKRVRLEMEGK 392
Db 503 MNVAINELNIKLODTNSQEKLEBELGLILEE---KC---ALQRLQEDL---VEELSFS 552
QY 393 DEEITKTEAQCRELKKLQEEHHHS---KELRLVEVEKLQKRMSLEKLEEA FRSKSEK 448
Db 553 REQIQRARQTIAEQESKL-NEAHKSLSTVEDLKAEIVSASESRKELELKHEA-----EV 605
QY 449 TQLHLN---LEKEXNLTKDLL---NELEVVKSRVKELECSERLEKAELSLKDDLTKLKS 502
Db 606 TNYKIKLEMLEKEKNAVLDRMAESQEAELERLTQLLFSHEEELSK---LKEDLE--- 657
QY 503 FTVMLVDERKNMMEKIKQE-----ERKVDGLNKNPKVEQGVMDVTEKLI EESKLLKLK 557
Db 658 -----IEHRIN-IEKLKDNLGIHYKQIDGLQNEWSQKIETMQFEKDNLITKQNQLILEI 711
QY 558 SEMEKEYSLTKERDELGMGLRSEERSCELSVDLLKKRLDGI EEVEREINRGRSCKG 617
Db 712 SKULDLOQQSLVNSKSE-----EMTLQINELQKEIETLRQEEKE-----KG 751
QY 618 SEFTCPEDNKIRELTLEIERLKKRLQOLE-----VVEGDLMKTEDEYDQLEQKFR 667
Db 752 T-----LEQEVQBELQKTELLEKQMKKEKENDLQEKFAQLEAENSILK--DEKKTLEDMLK 804
QY 668 -----TEQDKANFLS-----QQLEEIKHQMAK-HKAIEKGEAVSQE 702
Db 805 IHTPVSQEERLIFLDSIKSKSDSVWEKEIEIIEENEDLKQOCIQLNEEIEEKQNTFSF 864
QY 703 AE-----LRRFRLEEAKSR---DLQAEVQALKEKIH----- 731
```

Db 865 AEKNFEVNYQELQEEYACLLKVKDDLEDSSKNQKELEYKSKLKALNEELHLQRIINPTVKM 924
QY 732 -----ELMNKED-----QLSQLQVDYS-VLQQR--- 753
Db 925 KSSVFDEDEKTFVAETLEMGVEVVEKDTTELMKLEVTKREKLELSQRLSDLSEQLKQKHGE 984
QY 754 --FMEETKNKMGREVLNL-TKELELSKRYSRALRPSGNGRRMVDVPVAS--TGQVTEA 808
Db 985 ISFLNEEVKSLQKEQVSLRCRELEIIINHNR-----ENVQSCDTQVSSLLDGVVVTMT 1039
QY 809 VCGDAAEEETPAVFIRKSFQENHIM-----SNLRQVGLKKPMERSVLDVYPPAANE 861
Db 1040 SRG----AEGSVKVNKSFGEESKIMVEDKVSFENM-TVGEESKQEQ-LILDHLPSTVKE 1093
QY 862 LTMR 865
Db 1094 SSLR 1097

RESULT 10
US-08-714-741-41
; Sequence 41, Application US/08714741
; Patent No. 6500613
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yother, Janet
; APPLICANT: Crain, Marilyn J.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tart, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
; TITLE OF INVENTION: PORTIONS AND PRODUCTS
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.
; ZIP: 10036

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,741
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer Esq., William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454312-2460
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1231 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
US-08-714-741-41

Query Match 6.8%; Score 413.5; DB 4; Length 1231;
Best Local Similarity 21.1%; Pred. No. 1.2e-15;
Matches 215; Conservative 185; Mismatches 359; Indels 261; Gaps 42;

QY 135 EKSIGEDVYEKPISELDRLBEEKQKETYRRMLLEQLLAEKCHRRRTVYELENEKHKHTD--- 191
Db 2 DSSVGEETLPS--SLNMANESQTE-HRKDVDEYI-----KQMLSEIQLDRRKHTQNEE 52
QY 192 ----YMNKSDDFTNLLEQERERLKKLLEQEKAYQARKEKENAKRLNKLRLDELVKLSPAL 247
Db 53 SPVASQSKAEKDYDAAKDAKNAKKAVE-----DAQKALDDAKAAQKYDEDEVNL-NIKL 106
QY 248 MLVDERQMHIQGLQSQKVQDLTKQLREEEEEKLKAVTYKSKEDRQKLLKLEVDFEHKAS 307
Db 107 SAIKTKYLYELSVLKENSKEELTSKTAE---LTAAFEQFKD-----TLQKTEEKAA 158
QY 308 RFSQHEHEMN-----AKLANQESHNRQLRLKLVGLSQRIELEEETNKSLOKAEHELQ 359
Db 159 LEKAASEEMDKAQAQAYLAYQOATDKPEK-----KVAEAEKKVEEAKKAK 207
QY 360 ELREKIAKGECCGNSLMAEVESLRKRVLEMEGKDEEITKTEAQCRELKKLQBEHHHSKE 419
Db 208 DQKEEDRRNYPTNTYKTLLEIAESDVVKAAKD---AADKMIDEAKKREEEAKTKFNT 263
QY 420 LRLEVEKLQKRMSELEKLEBAFSSRSKSECTQLHLNLEKEKNLTKOLLNELEVVKSRVKEL 479
Db 264 VRAMVVK---EAELELVKEEANESRNE-----EKIKQAKEKVESK 300
QY 480 ECSESRLKAELSLKDDLTKLKSFTVMLVDERKNMMEKIKQEEKVGDGLNKNPKVEQGKV 539
Db 301 KAEATRLEK---IKTDRKKAEEPEPEQLAETKKKSEAKQ---KAPELTK--KLEAK- 350
QY 540 MDVTEKLIIESKKLLKLMSEEMEEKEYSLTKERBELMGKLRSEERSCELSVDLLKKRL 599
Db 351 ----RKAESESEKKAEEAKQVDAEYAL-----EAKIAELEVEYQVORLEKEL 392
QY 600 DGI-----EVEIREINRGRSCKGSEFTCPEDNKKIRELTLEIERLKKRLQLEVEVGD 651
Db 393 KEIDEEAKAKLEBAKKAATEAKQKVDAEVAPO-AKIAELENOVHRLEQELKEIDESDSE 451
QY 652 LMKTEDEYDQLEQKFRTEQ---DKANFLSQLEEEIKHOMAKHKA-IEKGEAVSQEAELRH 707
Db 452 DYLKEGLRAPLQSKLDTKKAKLSKLEELSDKIDELDAEIAKLEQLKDAEGNNNVEAYF 507
QY 708 RFRLEEAKSRDLQAEVQALKEKIHLMNKEQLSQLQVDYSVLQQRFMEEETKNK----- 762
Db 508 -FR-----APLQSKLDAKAKLSKLEELSDKIDELDAEIAKLEQLKDAEGNNNVEAYF 560
QY 763 NMGREVLNLTKELELSKRYSRALRPSGNGRRMVDVPVASTGVQTEAVCGDAAEEETPAVF 822
Db 561 KEGLEKTTAEKKAELEK-----AEADLKKAVIDEP-----ETPAPA 595
QY 823 IRKSFOEENHIMNLQVGLKKPMERSVLDVRYPPAANELTMRKSWIPWMRKRENGPSTP 882
Db 596 PQAAAEENNVEDYFKE-GLEKTI-----AAKKAELEKTEAD-LKKAVNEPEKP 642
QY 883 ---QEKGPRPNQAGHPGELVLAPKQGPPLHIRVTPDHENSTATLEITSPTSEFFSSTT 939
Db 643 APAPEPAPAEKPAEKP-----APAPEKP-----APAPEK----- 672
QY 940 VIPTLGNQKPRITIIPSPNVMSQKPKSADPTLGPERRAMSPVTITTISREK-----SPEGGR 995
Db 673 --PAPAEKP-APATPAPAEPAEQPK---PAPAPQAPAPAPKPEKPAEQPKPKTDQQAQAE 726
QY 996 SAFADRPASPIQIMTVSTSAAPTEIAVSPES-----QEVPMG----- 1032
Db 727 EDYARRPEKP-----APAPEKPAPTPTPTKGTGWKQENGWYFYNTDGSMTGWSEEE 778
QY 1033 -RTILKVTPEKQTVAPVRK-----YNSNANIITTEDNKIHLGCSQFKRSPG 1079
Db 779 YNRLTQQQPPKAEKPPAPAPKGTGWKQENGWYFYNTDGSGL---QNNGSWYILNSNGMATG 835

RESULT 11
US-09-310-187A-1
; Sequence 1, Application US/09310187A
; Patent No. 6358751

```
; GENERAL INFORMATION:
; APPLICANT: Benichou, Gilles
; APPLICANT: Fedoseyeva, Eugenia
; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
; TITLE OF INVENTION: Graft Rejection
; FILE REFERENCE: UCSF-090
; CURRENT APPLICATION NUMBER: US/09/310,187A
; CURRENT FILING DATE: 1999-05-12
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-310-187A-1

Query Match          6.7%; Score 409.5; DB 4; Length 1939;
Best Local Similarity 21.1%; Pred. No. 3.7e-15;
Matches 200; Conservative 181; Mismatches 297; Indels 269; Gaps 37;

Qy 35 DAKKNKANRK--EEDVMASTGIKRHLKPSGESEKTKTKSVLSK--EDLIQLLSIMEGE 89
Db 918 EAKVKEMNERLEDEEMNAELTAKK-----RKLEDECSSELKKDIDDLTLAKVEKE 969

Qy 90 LQAREDVIHMLRTE-----KTKPEVLEAHYGSABPEKVLRLVHRDAILAQEKS 137
Db 970 KHATENKVKNLTEEMAGLDLIIAKLTKEKKALQEAHQALDDLOV----BEDKVNLSKS 1025

Qy 138 IGEDVVEKPISELD-RLEEKQK-----ETVRRMLEQLLLAELKCHRRRTVYLENEKHKHTD 191
Db 1026 --KVKLEQVDDLEGSLEQEKVKRMDLERAKRKLEGLD---KLQTESIMDLENDKLQLEE 1080

Qy 192 YMNKSDFTNLLQEQUERERLKKLEQEKAQYQARKE-KENAKRLNKLRLDEL-----V 240
Db 1081 KLKKKEFDIN---QQNSKIED--EQALALQQLQKKLKENQARIEELEEELEAERTARAKVE 1135

Qy 241 KLKSFALMLVDROMHIEQLGLQSQKVDLTQKLREEEELKAVTYKSKEDRQKLLKLEV 300
Db 1136 KLRSDLRSLEEEISERLEEAGGATSVQIENMKKRAEAFQKM-----RRDLEEAATL 1185

Qy 301 DFEHKASRFSQEHEEMNAKLANQ-----ESHNRQLRLKLVGLSQRIEELEETNK 349
Db 1186 QHEATAAALRKKHADSVAEELGEQIDNLQRVKQKLEKEKSEFKLELDDVTNSMEEQIIKAKA 1245

Qy 350 SLQKA-----EELQELREKI-----AKGECNSSLMAEVESLR-----383
Db 1246 NLEKVSRTLEDQANEYRVKLEEAQRSNDFTTORAKLQTENGELARQLEKEALISQLTR 1305

Qy 384 -----KRVLEMEGKDEE-----ITKTEAQCRELKKKLQEEHHHSKELRL-----422
Db 1306 GKLSYTOQMEDLKRQLEEGKAKNALAHALQSAHRDCDLLREQYEEETEAKAELQRVLSK 1365

Qy 423 -----EVEKLQKRMSE-LEKLEBAFSSRSKSECTQLHLNLEKEKN 460
Db 1366 ANSEVAQWRTKYETDAIQRTTEELEEAKKKLAQRLQDAEEAVEAVNAKCSLSLEKTKHRLQN 1425

Qy 461 LTKDLLNELE-----VVKSRVKELECSERLE---KAELSLKDDLT 499
Db 1426 EIEDLMVDVRSNAAAAALDKKQNFDKILAEWKQKYEBESQSELESSQKEARSLSTELPK 1485

Qy 500 LKSFTVMLVDERKNMMEKITQOEERKVDGLNKNFKVQGVMDVTEKLIIESKKL-----553
Db 1486 LKN---AYEESLEHLETPKRE-----NKNL---QEBISDLTEQLGEGGKNVHELEKV 1531

Qy 554 -----LKLKSEMEEEKYSLTKERDELM-----GKLRSSEERSC-----586
Db 1532 RKQLEVEKLEQLQSALAEABASLEHEEGKILRAQLEFNQIKAEIERKLAEBKDEEMEQAKRN 1591

Qy 587 -----ELSCSVDL-----LKKRLDG-IEEVEREINGRSCSKGEFTCPEDNKIR 629
Db 1592 HQRVVDSLSQTSLDAETRSRNEVLRVKKKMEGDLNEMETQLSHA-----NRMAA 1639

Qy 630 ELTLEIERLKKRLOQLEVVEGDLMKTEDEYDQLEQKFRTEQDKANFLSQOLEEIK-----684
```

```
Db 1640 EAQKQVKSLSQLLKDTQIQIQLDDAVRANDD---LKENIAIVERRNLLQAELEELRAVVEQ 1696
Qy 685 ----HQMAKHKAIEKGEAV---SQEAELRHRFRLEEAKSRLDLQAEVQALKEKIHLMNK 736
Db 1697 TERSRKLAEQELIETSERVQLLHSQNTSLNQKKMESDLTLQLQSEVE---EAVQECRNA 1753

Qy 737 EDQLSQLQVDYSVLQRFEMEE-----ETKNKNMGREVLNLTKELE 776
Db 1754 EEKAKKAITDAAMMAEELKKEQDTSAHLERMKNKMEQTTIKDLQHRLD 1800

RESULT 12
US-08-938-105-3
; Sequence 3, Application US/08938105
; Patent No. 6353151
; GENERAL INFORMATION:
; APPLICANT: Leinwand, Leslie A.
; APPLICANT: Vikstrom, Karen L.
; TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,105
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crook, Wannell M.
; REGISTRATION NUMBER: 31,071
; REFERENCE/DOCKET NUMBER: 3595-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1886 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-938-105-3

Query Match          6.7%; Score 407.5; DB 4; Length 1886;
Best Local Similarity 22.7%; Pred. No. 4.6e-15;
Matches 219; Conservative 171; Mismatches 360; Indels 216; Gaps 40;

Qy 36 AKKNKANRKEEDVMASTGIKRHLKPSGESEKTKTKSVLSKEDLIQLLSIMEGELQARED 95
Db 919 ATENKVKNLTEEMAGLDLIIAKL-----TKEKKALQEAHQQALD-----DLQAEED 964

Qy 96 VIHMLRTEKTKPE-VLEAHYGSAPPEKVLRV-----LHRDAILAQEK-----S 137
Db 965 KVNTLTQSKVKVLEQQVDDLEGSLEQEKVKRMDLERAKRKLEGLDKLTQESIMDLENDKLQ 1024

Qy 138 IGEDVVEKPI-----SELD-----RLEEKQKETVRRM--LEQLLLAEKCHRRTVVEL 182
Db 1025 LEEKLKKKEFDISQQNSKIEDEQALALQQLKLENQARIEELEEELEAERTARAKVEKL 1084

Qy 183 ENEKHKHTDYMKNKSDFTNLLQEQUERLKKL-----LEQEKAYQAPKEKENAKRLNKL 235
Db 1085 RS-----DLTRELEEISERLEEAGGATSVQIEMNK----KREAEFQKMRRDL 1127

Qy 236 RDELVKLKSFALMLVDERQMHIQGLQSQKQVQDLTQKLREEEEKLKA-----VT-----285
```

Db 1128 BEATLQHEATAAALRKKHADSVAEELGEQIDNLRVQKLEKESEFKLELDDVTSHMEQI 1187
QY 286 YKSKEDRQKL-----LKLEV-----DPEHKASRFSQEHEEMNAKLANQESHN 327
Db 1188 IKAKANLEKVSRTLEDQANEYRVKLEEAQRSNDFTTQRAKLTQENGELARQLKEKEALI 1247
QY 328 RQLRLKLVGLSORIE-----ELEETNK-----SLQAAEEELQELREKIAKGECCNSSLM 376
Db 1248 WQLTRGKLSYTOQMEDLKQLEEEGKAKNALAHALQSAHRDCLLREQYEEEMAKAELQ 1307
QY 377 AVEVESLRKRVLEMEGKDE--EITKTEAQCRELKKLQEBEHHHSKELRLEVEKLOKRMSEL 434
Db 1308 RVLSKANSEVAQWRTKYETDAIQORTE-ELEEAKKLAQRLQDAEE---AVEAVNAKCSSL 1363
QY 435 EKLEEAFSRSKSECTQLHLNLE-----KEKNLT KDLLNELEVVKSRVKELECSE 483
Db 1364 EKTKH---RLQNEIEDLMVDVERSNAAAAAALDKQRNFDKILAEWKQKYEESQSELESS- 1419
QY 484 SRLEKAELSLKODLTCLKSFTVMLVDERKNNMMEKIQEERKVDGLNKNFKVEQGVMDVT 543
Db 1420 ---QKEARSLSTELFKLN---AYEESLEHLETFKRE-----NKNL---QEISDLT 1462
QY 544 EKLIEESK---KLLKLSMEEEKEYSLTKERDELGMKLRSEERSCELSVLLKKRLD 600
Db 1463 EQLGEGGKNVHELEKIRKQLEVEKLELQSALEEAESLEHEEGK-----ILRAQLE 1513
QY 601 GIEEVEREINRGRSCKSEFTCPEDNKIR-----ELTLEIE-----RLKKRLQOLEV 647
Db 1514 -FNQIKAEIERKLAEDDEMEQAKRNHLRVVDSLQTSLDATRSRNEALRVKKM----- 1567
QY 648 VEGDLMKTEDEYDQLEQKFRTEQDKANFLSQQLEETKHQMA--KHKAIEKGEAVSQEAEL 705
Db 1568 -EGDLNEMEIQLSQANR-----IASEAQHLKNAQAHLKDTQLQLDDAVRANDDL 1616
QY 706 RHRFRLEEAKSRDQAEVQALKEKIHLMNKKEDQLSQLQVDYSVLQQRFMEEET---KN 761
Db 1617 KENIAIVERNTLLQAELEELRAVVEQTERSRLKAEQELIETSERVQLLHQSNTSLINQK 1676
QY 762 KNMGREVLNLTKELELSKRYSRALRPSGNGRRMVDVPVASTGVQTEAVCGDAEEETPAV 821
Db 1677 KQMDADLSQLOTEVE-----EAVQECRNAEEAKKAITDAAMMAEEL---KKEQDTSAH 1727
QY 822 FIR-----KSFQEEHNHMSNLRQVGLKKPMERSSVLD-RYPPAANELTMRK-----SWIP 870
Db 1728 LERMKNMEQTIKDQHLRLDEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVK 1787
QY 871 WMRKRE 876
Db 1788 GMRKSE 1793

RESULT 13
US-08-466-390-4
; Sequence 4, Application US/08466390
; Patent No. 5686562
; GENERAL INFORMATION:
; APPLICANT: TOUKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,390
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-466-390-4

Query Match 6.7%; Score 407.5; DB 1; Length 2101;
Best Local Similarity 20.3%; Pred. No. 5.3e-15;
Matches 266; Conservative 228; Mismatches 524; Indels 292; Gaps 50;

QY 65 EKTTSKSVLSKEDLIQLLSIMEGELQAREDIHMLRTEKTKPEVLEAHYGS--AEPEKV 122
Db 907 EKWAATSKEVARLETIVRKAGEQQTASRELKVEKPARAGDRQPEWLEEQQGRQFCSTQAA 966
QY 123 LRVLHRDAILAQEKSIGEDVYEKPISELDRLEEKQKETYRRMLQLLLAEKCHRTTVYEL 182
Db 967 LQAMEREA-----EQMGNELRLRAALMESGQQQEE-----RGQQR 1004
QY 183 ENKHKHTDYNMKSDDFTNLLQEQRERLKKLLEQEKAYQARKEKENAKRLNKLRLDELVKL 242
Db 1005 EVAR-----LTQERGRAQADLALAKAAAELEMLRLQNALNEQRFATL 1048
QY 243 KSFALMLVDERQMHIEQL---GLSQKVODLTQKLREEEKLVKAVTYKSKED----- 291
Db 1049 QEALAHALTEKEGKDQELAKLRGLEAAQIKEL--EELRQTVKQLKEQLAKKEKESGSGA 1107
QY 292 -----RQKLLKLEVDFFEKASRFSQEHEEMNAKLANQESHNRQLRLK 333
Db 1108 QSEAAGRTEPTGPKLEALRAEVSKLEQCCQKQEQADSLERSLEAERASRAERDSALETL 1167
QY 334 LVGLSQRIEELIETNKSLSQAAEELQELREK-----IAKG-----ECGN--- 372
Db 1168 QGQLEEKAAQELGHSQSALASAQRELAAFRTKVQDHSKAEDWKVAQVARGQEAERKNSLI 1227
QY 373 SSLMAEVESLRKRVLEMEGKDEEI-----TKTEAQ----- 403
Db 1228 SSLEEEVSILNRQVLEKEGESKELRLVMAESEKSQLLESCACCRQRPATVPQLQNA 1287
QY 404 -----RELKKKLOEEHHHSKELRLEVEKLOKRMSELEKLEAFSRKSECTQLH 452
Db 1288 LLCGRRRCRASGREAEKQORVASENLRLQELTSSQAERAEELGQELKAWQKFFQKEQALSTLQ 1347
QY 453 LNLEKEKNLT KDLLNELEVVKSRVKELECSERLEKAELSLKDDLTCLKSFTVMLVDERK 512
Db 1348 L-----EHTSTQALVSELLPAKHLCCQQLQABQAAAEKRH----- 1381
QY 513 NMMEKIKQEEERKVDGLNKNF--KVEQGVMDVTVEKLIIESKLLKLSXEMEE--KEYSL 567
Db 1382 --REELEQSKQAAGGLRAELLRAQRELGELIPLRQKVAEQERTAQQQLRAEKASYAEQLSM 1439
QY 568 TKERDELMGKLRSEERSCELSVLLKKRLD-GIEEVEREINRGRSCKSEFT---CP 623
Db 1440 LKKAHGLL-----ABENRG-----LGERANLGRQFLEVLDOAREKYVQELAAVRAD 1486
QY 624 EDNKIRELTLEIERLKKRLQOLEV-VEGDLMKTEDEYDQLEQKFRTEQDKANFLSQLE 682
Db 1487 AETRLAEVQREAAQSTARELEVMTAKYEGAKVKVLEE---RQRFQEEERQK---LTAQVEE 1539
QY 683 IKHQMAKHAIEKGEAVSQEAELRHRFRLEEAKSRDLQAEVQALKEKIHLMNKKEDQLSQ 742

Db 1702 PQLDLSIDSLD-LSCEEGLTSLTSKLPRTQPDGTSVPGEPASPISORLPPKVESLESY 1760
Qy 913 VTPDHENSTATLEITSPTSEFF-----SSTVTIPTLGNQKPRITIIPSPNVM 960
Db 1761 FTPIPARSQAPLESSLDSGLDGVFLDSGRKTRSARRRTTQIINIITMTKKLDVEEPDSANSS 1820
Qy 961 SQPKSADPTLGPAMSPVTITISREKSPGGRSAFADRPASPIQIMTVSTSAAPTE 1019
Db 1821 FYSTRSAPASQASLRATS--STQSLARLGSPDYGNSALLSLPGYRP---TTRSSARRSQ 1874
Qy 1020 IAVSPESQEVPMGRTILKVTPKQTVPAVPRKYNNSNANIITTEDNKI---HIHLGSQFKR 1076
Db 1875 AGV---SSGAPPGRNSFYM-GTCQDEPEQLDDWNRIABL--QQRNRVCPPHLKTCYPLES 1928
Qy 1077 SPGPAAEGVS-----PVITVR-----PVNTAEKEVSTGTVLRSPRNHLS----- 1116
Db 1929 RPSLSLGTITDEEMKTGDPQETLRRASMQPIQI-AE---GTGITTRQQRKRVSLPEHQGP 1984
Qy 1117 SRPGASKVTSTI--TITP-----VTTSSTRGTQSVSGQDGSSQRP-----TPTR 1158
Db 1985 GTPESKKATSCFPRTPTDRRHEGRKQSTTEAQKKAAPASTKQADRRQSMAFSILNTPKK 2044
Qy 1159 IPMSKGMKAGKPVVAASGAGNLTKFPRAETQSMKIELKKSAAASSTASLG 1208
Db 2045 LGNSL-LRRGASKKALSASPNTRSGTR---RSPRIATTASATAAAIG 2090

RESULT 15
US-08-467-781-4
; Sequence 4, Application US/08467781
; Patent No. 5780596
; GENERAL INFORMATION:
; APPLICANT: TOUKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,781
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-467-781-4

Query Match 6.7%; Score 407.5; DB 1; Length 2101;
Best Local Similarity 20.3%; Pred. No. 5.3e-15;
Matches 266; Conservative 228; Mismatches 524; Indels 292; Gaps 50;

Qy 65 EKKTKKSVELSKEDLIQLLSIMEGELQAREDVIHMLRTEKTKPEVLEAHYGS--AEPKVV 122
Db 907 EKMAATSKVARLETIVRKAGEQOETASRELVKEPARAGDRQPEWLEEQGRQFCSTQAA 966
Qy 123 LRVLHRDAILAQEKSIGEDVYEKPISELDRLEEKQKETYRRMLEQLLLAEKCHRRTVYEL 182
Db 967 LQAMEREA-----EQMGNELRLRAALMESGQQQEE-----RGQQR 1004
Qy 183 ENKHKHTDYMKNKSDFTNLLEQERERLKKLEQKAYQARKEKENAKRLNKLDELVKL 242
Db 1005 EVAR-----LTQERGRAQADLALEKAARAELEMLQNALNEQRFVFPATL 1048
Qy 243 KSFALMLVDERQMHIQOL---GLQSQKQVDLTQKLEEEEEKLKAVTYKSKED----- 291
Db 1049 QEALAHALTEKEGKDQELAKLRGLEAAQIKEL--BELRQTVKQLKEQLAKKEKHEHSGSGA 1107
Qy 292 -----RQKLLKLEVDFEHKASRFSQEHEEMNAKLANQESHNRQLRLK 333
Db 1108 QSEAAGRTEPTGPKLEALRAEVSKLEQQCQKQEQADSLERSLEAERASRAERDSALETL 1167
Qy 334 LVGLSQRIELEEETNKSLOKAEELQELREK-----IAKG-----ECGN--- 372
Db 1168 QGQLEKAQELGHSQALASAQRELAAFRTKVQDHSKAEDWKQAQVARGRQEAERKNLSI 1227
Qy 373 SSLMAEVESLRKRVLMEGKDEEI-----TKTEAQC----- 403
Db 1228 SSLEEEVSILNRQVLEKEGESKELKRLVMAESEKSKLESCACCRCRQRPATVPBLQNA 1287
Qy 404 -----RELKKKLQEEHHHSKELRLEVEKIQKRMSELEKLEAFRSKSECTQLH 452
Db 1288 LLCRRRCRASGREAEKQRVASENLRQELTSQAERAEEELGQELKAWQEKFFQKEQALSTLQ 1347
Qy 453 LNLEKEKNLT KDLLNELEVVKSRVKELECSSESLEKAELSLKODLTCLKSFTVMLVDERK 512
Db 1348 L-----EHTSTQALVSELLPAKHLCCQQLQAEQAEEAKRH----- 1381
Qy 513 NMMEKIKOEERKVDGLNKNF---KVEQKQKMDVTEKLIIEESKLLKLKSEMEE--KEYSL 567
Db 1382 --REELEQSKQAAGGLRAELLRAQRELGELIPLRQKVAEQERTAQQLRAEKASYAEQLSM 1439
Qy 568 TKERDELMGKLRSSEERSCELSVDLLKKRLD-GIEEVEREINRGRSCKGSEFT---CP 623
Db 1440 LKKAHGLL----AEENRG-----LGERANLGRQFLEVELDQAREKYVQELAAVRAD 1486
Qy 624 EDNKIRELTLEIRLKKRLQQLV-VEGDLMKTEDEYDQLEQKFRTEQDKANFLSQLEE 682
Db 1487 AETRLAEVQREAOASTARELEVMTAKYEGAKVKVLEE---RQRFQEEERQK---LTAQVEE 1539
Qy 683 IKHQMAKHKAIEKGEAVSQEAELRHRFRLEEAASRDLOAEVQALKEKIHLMNKEDQLSQ 742
Db 1540 LSKKLADS---DQASKVQQQ-----KLKAVQAQGGESQOEAFQFQAQLNELQAQLSQKEQ 1591
Qy 743 LQVDYSVLQOREFNEEETKNMGMREVNLTKLELELSKRYSRALRPSGNGRRMVDVVPVAST 802
Db 1592 AAHYKLQMEKAKATHYDAKKQQNQELQELRSLEQLQKENKELRAEA---ERLGHQLQA 1648
Qy 803 GVQT---EAVCGDAAEEETPAVFIKRSFQENHIMSNLRQVGLKKPMERSVLDRIYPPAA 859
Db 1649 GLKTKEAEQTCRHLTAQ-----VRSLEAQVAHADQQLRDLG-KFQVATDAKLSREPQAK 1701
Qy 860 NELTMRKSWIPWMRKRENGPSTPQEKGPRPN-QGAGHPGELV-----LAPKQGPPLHIR 912
Db 1702 PQLDLSIDSLD-LSCEEGLTSLTSKLPRTQPDGTSVPGEPASPISQRLPPKVESLESY 1760
Qy 913 VTPDHENSTATLEITSPTSEFF-----SSTVTIPTLGNQKPRITIIPSPNVM 960
Db 1761 FTPIPARSQAPLESSLDSGLDGVFLDSGRKTRSARRRTTQIINIITMTKKLDVEEPDSANSS 1820
Qy 961 SQPKSADPTLGPAMSPVTITISREKSPGGRSAFADRPASPIQIMTVSTSAAPTE 1019
Db 1821 FYSTRSAPASQASLRATS--STQSLARLGSPDYGNSALLSLPGYRP---TTRSSARRSQ 1874
Qy 1020 IAVSPESQEVPMGRTILKVTPKQTVPAVPRKYNNSNANIITTEDNKI---HIHLGSQFKR 1076

Db 1875 AGV---SSGAPPGRNSFYM-GTCODEPEQLDDWNRIAEI--QQRNRVCPPHLKTCYPLES 1928
QY 1077 SPGPAAEGVS-----PVITVR-----PVNVTAEKEVSTGTVLRSPRNHLS----- 1116
Db 1929 RPSLSLGTITDEEMKTGDPQETLRRASMOPIQI-AE---GTGITTRQQRKRVSLEPHQGP 1984
QY 1117 SRPGASKVTSTI--TITP-----VTTSTRGTQSVSGQDGSSQRP-----TPTR 1158
Db 1985 GTPESKKATSCFFPRPMTPRDRHEGRKQSTTEAQKKAAPASTKQADRRQSMAFSILNTPKK 2044
QY 1159 IPMSKGMKAGKPVVAASGAGNLTQFPRAETQSMKIELKKSAASTASLG 1208
Db 2045 LGNSL-LRRGASKKALSKASPNTRSGTR---RSPRIATTTAATAAAIG 2090

Search completed: September 7, 2004, 13:18:12
Job time : 45 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2004, 13:16:39 ; Search time 158 Seconds
(without alignments)
2416.829 Million cell updates/sec

Title: US-10-788-793-2
Perfect score: 6082
Sequence: 1 MRSRNQGGESSNGHVSCP.....KIELKKSAASTASLGGKG 1212

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1298764 seqs, 315065143 residues

Total number of hits satisfying chosen parameters: 1298764

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications_AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
 - 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
 - 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
 - 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
 - 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
 - 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
 - 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
 - 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
 - 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
 - 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
 - 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5301	87.2	1114	16	US-10-408-765A-2119 Sequence 2119, Ap
2	2452.5	40.3	1133	14	US-10-309-851-14 Sequence 14, Appl
3	2444.5	40.2	1133	14	US-10-309-851-16 Sequence 16, Appl
4	2444.5	40.2	1135	14	US-10-309-851-12 Sequence 12, Appl
5	2406.5	39.6	512	15	US-10-108-260A-4080 Sequence 4080, Ap
6	1828	30.1	764	14	US-10-309-851-10 Sequence 10, Appl
7	1814	29.8	824	15	US-10-108-260A-3569 Sequence 3569, Ap
8	1675	27.5	785	14	US-10-309-851-24 Sequence 24, Appl
9	1645.5	27.1	752	14	US-10-171-311-48 Sequence 48, Appl
10	780.5	12.8	439	9	US-09-925-302-495 Sequence 495, App
11	780.5	12.8	439	12	US-09-925-302-495 Sequence 495, App
12	769.5	12.7	350	14	US-10-309-851-8 Sequence 8, Appli
13	677	11.1	265	14	US-10-309-851-6 Sequence 6, Appli
14	618.5	10.2	384	14	US-10-309-851-38 Sequence 38, Appl
15	618.5	10.2	386	14	US-10-309-851-22 Sequence 22, Appl

16	602	9.9	274	15	US-10-104-047-3353	Sequence 3353, Ap
17	526.5	8.7	240	14	US-10-309-851-36	Sequence 36, Appl
18	518.5	8.5	240	14	US-10-309-851-4	Sequence 4, Appli
19	469.5	7.7	1961	15	US-10-028-248A-103	Sequence 103, App
20	469.5	7.7	1961	15	US-10-107-782-103	Sequence 103, App
21	469	7.7	1960	15	US-10-236-031B-62	Sequence 62, Appl
22	469	7.7	1960	15	US-10-028-248A-104	Sequence 104, App
23	469	7.7	1960	15	US-10-107-782-104	Sequence 104, App
24	468.5	7.7	1959	15	US-10-028-248A-36	Sequence 36, Appl
25	468.5	7.7	1959	15	US-10-107-782-36	Sequence 36, Appl
26	464.5	7.6	1790	15	US-10-369-493-1586	Sequence 1586, Ap
27	463.5	7.6	2020	15	US-10-369-493-5128	Sequence 5128, Ap
28	463.5	7.6	2020	15	US-10-369-493-5129	Sequence 5129, Ap
29	461.5	7.6	1881	14	US-10-032-585-7646	Sequence 7646, Ap
30	461.5	7.6	1999	15	US-10-028-248A-107	Sequence 107, App
31	461.5	7.6	1999	15	US-10-107-782-107	Sequence 107, App
32	458.5	7.5	1961	15	US-10-028-248A-105	Sequence 105, App
33	458.5	7.5	1961	15	US-10-107-782-105	Sequence 105, App
34	457	7.5	1164	15	US-10-369-493-6564	Sequence 6564, Ap
35	452	7.4	1938	14	US-10-171-311-164	Sequence 164, App
36	452	7.4	1945	10	US-09-927-597-2	Sequence 2, Appli
37	452	7.4	1972	14	US-10-171-311-162	Sequence 162, App
38	452	7.4	1972	15	US-10-341-434-103	Sequence 103, App
39	452	7.4	1979	10	US-09-927-597-4	Sequence 4, Appli
40	451	7.4	1905	15	US-10-259-194A-86	Sequence 86, Appl
41	447.5	7.4	1959	15	US-10-028-248A-106	Sequence 106, App
42	447.5	7.4	1959	15	US-10-107-782-106	Sequence 106, App
43	447	7.3	1978	15	US-10-094-466-64	Sequence 64, Appl
44	441	7.3	880	15	US-10-369-493-21643	Sequence 21643, A
45	439	7.2	1827	15	US-10-369-493-5368	Sequence 5368, Ap

ALIGNMENTS

RESULT 1

US-10-408-765A-2119
; Sequence 2119, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2119
; LENGTH: 1114
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2119

Query Match 87.2%; Score 5301; DB 16; Length 1114;
Best Local Similarity 94.6%; Pred. No. 3.5e-279;
Matches 1054; Conservative 28; Mismatches 32; Indels 0; Gaps 0;

QY 99 MLRTEKTKPEVLEAHYGSAPPEKVLRLVLRDAILAQEKSIGEDVYEKPISELDRLEEKQK 158

Db 1 MLKTEKTKPEVLEAHYGSAPPEKVLRLVLRDAILAQEKSIGEDVYEKPISELDRLEEKQK 60

QY 159 ETYRRMLEQLLLAEKCHRRRTVYELENEKHKHTDYMNKSDFTNLLEQERERLKKLLEQEK 218

Db 61 ETYRRMLEQLLLAEKCHRRRTVYELENEKHKHTDYMNKSDFTNLLEQERERLKKLLEQEK 120

QY 219 AYQARKEKENAKRLNKLRLDELVKLSFALMLVDERQMHIQLGLQSQKVQDLTQKLREEE 278

Db	768	TKELERYRHFSKSLRPSLNGRRISDPQVFSKEVQTEAV-----DNEPPDYKSLIPLERAV	822
QY	822	FIRKSFOE-ENHIMSNLRQVGLKKPMERSSVL-----DRYPPAANELTMRKSWIPWMRKR	875
Db	823	INGQLYEESQD-----EDPNDEGSVLSFKCSQSTPCPVN-----RKLWIPWMKSK	869
QY	876	E-----NGPSTPQEKGRPNQAGHPGELVLAPKQGQPLHIRVTPDHENSTATLEITSPTS	931
Db	870	EGHLQNG-----KMQTKPNANFVQPGDLVLSHTPGQPLHIKVTDPDHVQNTATLEITSPTT	924
QY	932	E--EFFSSTTVIPLGNQKPRITIIPSPNVMSQKPK-SADPTLGPERRAMSPVTITTISRE	988
Db	925	ESPHSYTSTAVIPNCGTPKQRITILQNASITPVKSKTSTEDLMNLEQGMSPITMATFARA	984
QY	989	KSPGGRSAFADRPASPIQIMTVSTSAAPTEIAVSPESQEVPMGRTILKVTPEKQTV PAP	1048
Db	985	QTPESCGSLTPERTMSPIOVLAVTGSASSPEQGRSPEPTEISAKHAIFRVSPDRQSSWQF	1044
QY	1049	VRKYNSNANIITTEDNKIHIHLGSOQFKRSPGPAAGVSPVITVRPVNVTTAEKEVSTGTVL	1108
Db	1045	QRNSNSNSSSVITTEDNKIHIHLGSPYMQA---VASPVRPASPSAPLQDNRTQGLINGALN	1101
QY	1109	RSPRNHLSSRPGASKVTSTITITPVTTSSSTRGTQ	1142
Db	1102	KT-----TNKVTSSTITITPTATPLPRQSQ	1125
RESULT 3			
US-10-309-851-16			
; Sequence 16, Application US/10309851			
; Publication No. US20030108554A1			
; GENERAL INFORMATION:			
; APPLICANT: Saus, Juan			
; APPLICANT: Revert-Ros, Francisco			
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity			
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein			
; FILE REFERENCE: 98,723-F-US			
; CURRENT APPLICATION NUMBER: US/10/309,851			
; CURRENT FILING DATE: 2002-12-04			
; NUMBER OF SEQ ID NOS: 38			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 16			
; LENGTH: 1133			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-10-309-851-16			
Query Match 40.2%; Score 2444.5; DB 14; Length 1133;			
Best Local Similarity 46.1%; Pred. No. 3.9e-124;			
Matches 541; Conservative 214; Mismatches 338; Indels 81; Gaps 19;			
QY	1	MRSRNQGGESSNGHVSCPKSSIISSDGGKPSDAKKNKANRKEE-DVMASGTIKRHLK	59
Db	1	MRSR--GSDTEGSAOKKFPRHT-----KGHSFQGPKNMKHRQQDKDSPESDV---IL	48
QY	60	PSGESEK----KTKKSVELSKEDLIQLLSIMEGELQAREDVIHMLRTEKTKPEVLEAHYG	115
Db	49	PCPKAEKPHSGNGHQAEDLSRDDLLFLLSILEGELQARDEVIGILKAEKMDLALLEAQYG	108
QY	116	SAPEKVLRLHRDAILAQEKSIGEDVYEKPISELDRLEEKQKETYRRMLEBQLLLAEKCH	175
Db	109	FVTPKKVLEALQRDAFQAKSTPWQEDIYEKPMNELDKVVEKHKESYRRILGQLLVAEKSR	168
QY	176	RRTVYELENEKHKHTDYMNKSDDFTNLLEQERERLKLLEQEKAYQARKEKENAKRLNKL	235
Db	169	RQTILEEEERKHKEYMEKSDEFICLLEQECERLKLIDQEIKSQEEKEQEKERVTTL	228
QY	236	RDELVKLKSFALMLVDERQMHIEQLGLSQKVQDLTQKLREEEEEKLKAVTYKSKEDRQKL	295
Db	229	KEELTKLKSFALMVVDEQQRLTAQLTLQRKIQIELTNAKETHHTKLALAEARVQEEEQKA	288
QY	296	LKLEVDFEHKASRFSQHEEMNAKLANQESHNRQLRLKLVGLSQRIEELEETNKSLOKAE	355

Db	289	TRLEKELQTTKFHQDQDTIMAKLTNEDSQNRQLQOKLAALSQRQIDELEETNRSLRKAE	348
QY	356	EELQELREKIAKCEGNSSLMAEVESLRKRVLMEGKDEEITTKTEAQCBELKKKLOEEEH	415
Db	349	EELQDIKEKISKGEYGNAGIMAEVEELRKRVLDMEGKDEELIKMEEQCQCDLKNKRLERETL	408
QY	416	HSKELRLEVEKLRQRMSELEKLEEAFFSRKSECTQLHLNLEKEKNLTKDLLNELEVVKSR	475
Db	409	QSKDFKLEVEKLSKRIMALEKLEDAFNKSKQECYSYLCNLEKERMTTKQLSQUELESKVR	468
QY	476	VKELECSSESRLKAEKLSKDDLTCLKSFTVMLVDERKNMMEKIKQEERKVDGLNKNFKVE	535
Db	469	IKELEAIESRLEKTEFTLKEDLTCLKTLTVMFVDERKTMSEKLIKKTEDKIQAASSQLQVE	528
QY	536	QGVMDVTEKLIIEESKLLKLKSEMEEEKYSLTKERDELMGKLRSEERSCELSCSVDLL	595
Db	529	QNKVTTVTEKLIIEETKRALKSKTDVEEKMYSVTKERDDDLKNKLKAEBEKGNDLLSRVNL	588
QY	596	KKRLDGIIEVEREINRGR----SCKGSEFTCPEDNKIRELTLEIERLKKRLQQLVEVEGD	651
Db	589	KNRLQSLEAIEKDFLKNKLNQDSGKSTTALHQENNNKIKELSQEVERLKLKDKMKAIEDD	648
QY	652	LMKTEDEYDQLEQKFRTEQDKANFLSQOLEEIKHQMAKHAIEKGEAVSQEAELRHRFRL	711
Db	649	LMKTEDEYETLERRRYANERDKAQFLSKELEHVVMELAKYKLAEKTE-TSHEQWLFKRLQE	707
QY	712	EEAKSRDLQAEVQALKEKIHLMNKEDQLSQLQVDYSVLQQRFMEEETKNKNMGREVLNL	771
Db	708	EEAKSGHLSREVDALKEKIHVMATEDLICHQGDHSLVQKKLNQENNRNLDLGREIENL	767
QY	772	TKELELSKRYSRALRPSNGRRMVDVPVASTGVQTEAVCGDAAEETP-----AV	821
Db	768	TKELERYRHFSKSLRPSLNGRRISDPQVFSKEVQTEAV-----DNEPPDYKSLIPLERAV	822
QY	822	FIRKSFOE-ENHIMSNLRQVGLKKPMERSSVL-----DRYPPAANELTMRKSWIPWMRKR	875
Db	823	INGQLYEESQD-----EDPNDEGSVLSFKCSQSTPCPVN-----RKLWIPWMKSK	869
QY	876	E---NGPSTPQEKGRPNQAGHPGELVLAPKQGQPLHIRVTPDHENSTATLEITSPTS	931
Db	870	EGHLQNG-----KMQTKPNANFVQPGDLVLSHTPGQPLHIKVTDPDHVQNTATLEITSPTT	924
QY	932	E--EFFSSTTVIPLGNQKPRITIIPSPNVMSQKPK-SADPTLGPERRAMSPVTITTISRE	988
Db	925	ESPHSYTSTAVIPNCGTPKQRITILQNASITPVKSKTSTEDLMNLEQGMSPITMATFARA	984
QY	989	KSPGGRSAFADRPASPIQIMTVSTSAAPTEIAVSPESQEVPMGRTILKVTPEKQTV PAP	1048
Db	985	QTPESCGSLTPERTMSPIOVLAVTGSASSPEQGRSPEPTEISAKHAIFRVSPDRQSSWQF	1044
QY	1049	VRKYNSNANIITTEDNKIHIHLGSOQFKRSPGPAAGVSPVITVRPVNVTTAEKEVSTGTVL	1108
Db	1045	QRNSNSNSSSVITTEDNKIHIHLGSPYMQA---VASPVRPASPSAPLQDNRTQGLINGALN	1101
QY	1109	RSPRNHLSSRPGASKVTSTITITPVTTSSSTRGTQ	1142
Db	1102	KT-----TNKVTSSTITITPTATPLPRQSQ	1125

RESULT 4

US-10-309-851-12
; Sequence 12, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
; CURRENT APPLICATION NUMBER: US/10/309,851
; CURRENT FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1

Db 481 ECSESRLKAEKSLKODLTKLKSFVTMLVDER 512

|||||

RESULT 6

US-10-309-851-10

; Sequence 10, Application US/10309851

; Publication No. US20030108554A1

; GENERAL INFORMATION:

; APPLICANT: Saus, Juan

; APPLICANT: Revert-Ros, Francisco

; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity

; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein

; FILE REFERENCE: 98,723-F-US

; CURRENT APPLICATION NUMBER: US/10/309,851

; CURRENT FILING DATE: 2002-12-04

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 10

; LENGTH: 764

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-309-851-10

Query Match 30.1%; Score 1828; DB 14; Length 764;

Best Local Similarity 50.7%; Pred. No. 6.7e-91;

Matches 386; Conservative 144; Mismatches 210; Indels 22; Gaps 7;

QY 1 MRSRNOGGESSNGHVSCPSSIISSDGGKGPSEDAKNKANRKEE-DVMASGTIKRHLK 59

|||||

Db 1 MRSR--GSDTEGSAQKFPRT-----KGHSFQGPKNMKHRQDKDSPESDV---IL 48

60 PSGESEK---KTKKSVELSKEDLIQLLSIMEGELQAREDDVIHMLRTEKTKPEVLEAHYG 115

49 PCPKAEKPHSGNGHQAEEDLSRDDLLFLLSILEGELQARDEVIGILKAEKMDLALLEAQYG 108

116 SAEPEKVLRLHRDALLAQEKSIGEDVYEKPISELDRLEEKQKETYRRMLEQLLAEKCH 175

109 FVTPKKVLEALQORDAQAKSTPWQEDIYEKPMNELDKVVEKHKESYRRILGQLLVAEKS 168

176 RRTVYELENEKHHTDYMNKSDDFTNLLEQERERLKKLLEQEKAYQARKEKENAKRLNKL 235

169 RQTILEEERKHKHEKMEKSDDEFICLLEQECERLKLIDQEIKSQEEKEQEKRVTTL 228

236 RDELVKLSFALMLVDERQMHIEQLGLSQKVQDLTQKLREEEKLVKAVTKSKEDRQKL 295

229 KEELTKLSFALMVVDEQQRLTQALTQKIQELTNAKETHTKLALAEARVQEEQKA 288

296 LKLEVDFEHKASRFSGHEEMNAKLANQESHNRQLRLKLVGLSQRISLEEETNKSLOKAE 355

289 TRLEKELQTQTTKFHQDDQTIMAKLTNEDSQNRQLQOKLAALSROIDELEETNRLKAE 348

356 EELQELREKIAKGECSNSSLMAEVESLRKRVLEMEGKDEEITKTEAQCRELKKKLQEEEH 415

349 EELQDIKEKISKGEYGNAGIMAEVEELRKRVLDMEGKDEELIKMEEQCRDLNKRLERTL 408

416 HSKELRLEVEKLQKRNSELEKLEEAFAFSRSKSECTQLHLNLEKEKNLTKDLLNELEVVKSR 475

409 QSKDFKLEVEKLSKRIMALEKLEDAFNKSKQECYSLKCNLEKERMVTTKQLSQELESKVR 468

476 VKELECSERLEKAEKLSKDDLTKLKSFTVMLVDERKNMMEKIKQEEKRVKVDGLNKNFKVE 535

469 IKELEAIESRLEKTEFTLKEDLTKLKTLTMFVDERKTMSEKLTCTEDKLQAASSQLQVE 528

536 QGKVMVDVTEKLIIEESKLLKLKSEMEEKEYSLTKERDELGMGLRSEEEERSCELSVDLL 595

529 QNKVTTVTEKLIIEETKRALKSKTDVEEKMYSVTKERDDLKNKLKAEEEKGNDLLSRVNL 588

596 KKRLDGIIEEVEREINRGR----SCKGSEFTCPEDNKIRELTLEIERLKKRLQQLQEVVEGD 651

589 KNRLQSLEAIEKDFLKNKLNQDSGKSTTALHOENNKIKELSQEVERLKLKDKMKAIEDD 648

652 LMKTEDEYDQLEQKFRTEQDKANFLSQOLEEIKHQMAKHAIEKGEAVSQEAELRHRFRL 711

Db 649 LMKTEDEYETLERRYANERDKAQFLSKELHVKMELAKYKLAEKTE-TSHEQWLFKRLQE 707

712 EEAKSRDLQAEVQALKEKIHLMNKKEDQLSQLQVDYSVLQQR 753

708 EEAKSGHLSREVDALKKEIHEYMATEDLICHLOGDHSVLQKK 749

RESULT 7

US-10-108-260A-3569

; Sequence 3569, Application US/10108260A

; Publication No. US20040005560A1

; GENERAL INFORMATION:

; APPLICANT: HELIX RESEARCH INSTITUTE

; TITLE OF INVENTION: No. US20040005560A1el full length cdna

; FILE REFERENCE: H1-A0106

; CURRENT APPLICATION NUMBER: US/10/108,260A

; CURRENT FILING DATE: 2002-03-27

; NUMBER OF SEQ ID NOS: 5458

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3569

; LENGTH: 824

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-108-260A-3569

Query Match 29.8%; Score 1814; DB 15; Length 824;

Best Local Similarity 47.1%; Pred. No. 4.3e-90;

Matches 401; Conservative 150; Mismatches 237; Indels 64; Gaps 14;

QY 318 AKLANQESHNRQLRLKLVGLSQRISLEEETNKSLOKAEELQELREKIAKGECSNSSLMA 377

2 AKLTNEDSQNRQLQOKLAALSROIDELEETNRLKAEELQDIKEKISKGEYGNAGIMA 61

378 EVESLRKRVLEMEGKDEEITKTEAQCRELKKKLQEEHHHSKELRLEVEKLQKRNSELEKL 437

62 EVEELRKRVLDMEGKDEELIKMEEQCRDLNKRLERETQSKDFKLEVEKLSKRIMALEKL 121

438 EEAFAFSRSKSECTQLHLNLEKEKNLTKDLLNELEVVKSRVKELECSERLEKAEKLSKDDL 497

122 EDAFNKSKQECYSLKCNLEKERMVTTKQLSQELESKVRIKELEAIESRLEKTEFTLKEDL 181

498 TKLKSFTVMLVDERKNMMEKIKQEEKRVKVDGLNKNFKVEQGKVMVDVTEKLIBESKLLKLK 557

182 TKLKTTLTMFVDERKTMSEKLTCTEDKLQAASSQLQVEQNKVTTVTEKLIBETKRALKSK 241

558 SEMEKEYSLTKERDELGMGLRSEEEERSCELSVDLLKKRLDGIIEEVEREINRGR---- 613

242 TDVEEKMYSVTKERDDLKNKLKAEEEKGNDLLSRVNLKNRLQSLAEIAIEKDFLKNKLNQD 301

614 SCKGSEFTCPEDNKIRELTLEIERLKKRLQQLQEVVEGDLMKTEDEYDQLEQKFRTEQDKA 673

302 SGKSTTALHQENNKIKELSQEVERLKLKDKMKAIEDDLMKTEDEYETLERRYANERDKA 361

674 NFLSQOLEEIKHQMAKHAIEKGEAVSQEAELRHRFRLEEAKSRDLQAEVQALKEKIHKL 733

362 QFLSKELEHVKNMELAKYKLAEKTE-TSHEQWLFKRLQEEEAESGHSLSREVDALKKEIHEY 420

734 MNKEDQLSQLQVDYSVLQQRFMEEETKNKNMGREVLNLTKELELSKRYSRALRPSGNRR 793

421 MATEDLICHLOGDHSVLQKLNQENNRNDRGREIENLTKELEYRHRHFSKLSRPSLNGRR 480

794 MVDVPVASTGVQTEAVCGDAEEETP-----AVFIRKSFQE-ENHIMSNLRQVGL 842

481 ISDPQVFSKEVQTEAV-----DNEPPDYKSLIPLERAVINGQLYEESENQD----- 526

843 KKPMERSSVL-----DRYPPAANELTMKRSWIPWPKRE----NGPSTPQEKGPRPNQGA 893

527 EDPNDEGSVLSPKCSQSTPCPVN-----RKLWIPWMKSKEGHLONG-----KMQTRPNANF 577

894 GHPGELVLAPKQGPPLHVRVTPDHENSTATLEITSPTSE--EFFSSTTVIPTLGNQKPRI 951

578 VQPGDLVLSHTPGQPLHIKVTPDHVQNTATLEITSPTTESPHSYTSTAVIPNCGTPKQRI 637

Db 121 GEYGNAGIMAEV-----EELIKMEEQCRDLNKRLERETLOSKDFKLEVEKL 166

QY 428 QKRMSELEKLEEAFFRSKSECTQLHLNLEKEKNLT KDLLNELEVKS RVKELECSERLE 487

Db 167 SKRIMALEKLEDAFNKSQECYSCLKNLEKERM TTKQLSQELES LKVRIKELEAESRLE 226

QY 488 KAE LSLKODLT KLKSF T VMLVDERKNMMEKIKOEERKVDGLNKNFKVEQGKVM DVTEKLI 547

Db 227 KTEFTLKEDLT KLKTLTVMFVDERKTMSEKLKKTEDKLQAASSQLQVEQNKVTVTEKLI 286

QY 548 EESKLLKLKSEMEKEYS LTKERDELGMGLRSEERSCELSV D L LKRLDGI EEVER 607

Db 287 EETKRALKSKTDVEEKMYSVTKERDDLKNKLKAE E BKGNDLLSRVNM LKNRLQSLEAIEK 346

QY 608 EINRGR---SCKGSEFTCPEDN KIRELTLEIERLKKRLQQLLEVVEGDLMKTEDEYDQLE 663

Db 347 DFLKNKLNQDSGKSTTALHQENNKIKELSQEVE RLKLKLDKMAIEDDLMKTEDEYETLE 406

QY 664 QKFRTEQDKANFLSQOLEEIKHQMAKHKAIEKGEAVSQEAELRHRFRLEEA KSRDLQAEV 723

Db 407 RRYANERDKAQFLSKELEHVKMELAKYKLAEKTE-TSHEQWLFKRLQEEEA KSGHLSREV 465

QY 724 QALKEKIH ELMNKEDQLSQLQVDYSVLQORFMEETKNKMGRVNLNLTKELELSKRYSR 783

Db 466 DALKEKIH EYMATEDLICH LQGDHSVCCKKLNQOENNRDLGREIENLTKELERYRHFSK 525

QY 784 ALRPSGNRRMVDVPVASTGVQTEAVCGDAAEETP-----AVFIRKSFQE-ENH 832

Db 526 SLRPSLNGRRISDPQVFSKEVQTEAV-----DNEPPDYKSLIPLERAVINGQLYEESENQ 580

QY 833 IMSNLRQVGLKKPMERSVL-----DRYPPAANELTMRKSWIPWMRKRE-----NGPSTPQ 883

Db 581 D-----EDPNDEGSVLSFKCSQSTPCPVN-----RKLWIPW M KSK EGH LQNG----- 622

QY 884 EKGPRPNQAGHPGELVLAPKQGOPLHIRVTPDHNSTATLEITSPTSE--EFFSSTTVI 941

Db 623 KMQTKPNANFVQPGDLVLSHTPGQPLHIKVTDPDHVQNTATLEITSPTTESPHSYTSTAVI 682

QY 942 PTLGNQKPRITIPSPNVMSQKPK-SADPTLGPERAMSPVTITTSISREKSP EGGRS AFAD 1000

Db 683 PNCGTPKQRITILQNASITPVKSKTSTEDLMNLEQGMSPITMATFARAQTPESCGSLTPE 742

QY 1001 RPAS 1004

Db 743 RTMS 746

RESULT 10

US-09-925-302-495

; Sequence 495, Application US/09925302

; Patent No. US20020044941A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PA104

; CURRENT APPLICATION NUMBER: US/09/925,302

; PRIOR FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05918

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

; PRIOR FILING DATE: 1999-03-12

; NUMBER OF SEQ ID NOS: 896

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 495

; LENGTH: 439

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (8)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: SITE

; LOCATION: (416)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-302-495

Query Match 12.8%; Score 780.5; DB 9; Length 439;

Best Local Similarity 41.7%; Pred. No. 2.3e-34;

Matches 186; Conservative 74; Mismatches 139; Indels 47; Gaps 11;

QY 665 KFRTEQDKANFLSQOLEEIKHQMAKHKAIEKGEAVSQEAELRHRFRLEEA KSRDLQAEVQ 724

Db 5 RYAXERDKAQFLSKELEHVKMELAKYKLAEKTE-TSHEQWLFKRLQEEEA KSGHLSREV 63

QY 725 ALKEKIH ELMNKEDQLSQLQVDYSVLQORFMEETKNKMGRVNLNLTKELELSKRYSR 784

Db 64 ALKEKIH EYMATEDLICH LQGDHSVLQKLNQOENNRDLGREIENLTKELERYRHFSKS 123

QY 785 LRPSGNRRMVDVPVASTGVQTEAVCGDAAEETP-----AVFIRKSFQE-ENHI 833

Db 124 LRPSLNGRRISDPQVFSKEVQTEAV-----DNEPPDYKSLIPLERAVINGQLYEESENQ 178

QY 834 MSNLRQVGLKKPMERSVL-----DRYPPAANELTMRKSWIPWMRKRE-----NGPSTPQE 884

Db 179 -----EDPNDEGSVLSFKCSQSTPCPVN-----RKLWIPW M KSK EGH LQNG-----K 220

QY 885 KGPRPNQAGHPGELVLAPKQGOPLHIRVTPDHNSTATLEITSPTSE--EFFSSTTVIP 942

Db 221 MQTKPNANFVQPGDLVLSHTPGQPLHIKVTDPDHVQNTATLEITSPTTESPHSYTSTAVIP 280

QY 943 TLGNQKPRITIPSPNVMSQKPK-SADPTLGPERAMSPVTITTSISREKSP EGGRS AFADR 1001

Db 281 NCGTPKQRITILQNASITPVKSKTSTEDLMNLEQGMSPITMATFARAQTPESCGSLTPER 340

QY 1002 PASPIQIMTVSTSAAPTEIAVSPESQEVPMGRITLKVTPKQTVPAVPVRKYNSNANIITT 1061

Db 341 TMSPIQVLAVTGSASSPEQGRSPEPTISA KHAIFRVSPDRQSSWQFORSNSNSSSVITT 400

QY 1062 EDNKIHILGSGQPKRSPGAAEGVSP 1087

Db 401 EDNKIHILGSPYMQXVASPVRPASP 426

RESULT 11

US-09-925-302-495

; Sequence 495, Application US/09925302

; Publication No. US20030064072A9

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PA104

; CURRENT APPLICATION NUMBER: US/09/925,302

; PRIOR FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05918

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

; PRIOR FILING DATE: 1999-03-12

; NUMBER OF SEQ ID NOS: 896

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 495

; LENGTH: 439

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (8)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: SITE

; LOCATION: (416)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-302-495

Query Match 12.8%; Score 780.5; DB 12; Length 439;

Best Local Similarity 41.7%; Pred. No. 2.3e-34;

Matches 186; Conservative 74; Mismatches 139; Indels 47; Gaps 11;


```

; LENGTH: 384
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-851-38

Query Match      10.2%; Score 618.5; DB 14; Length 384;
Best Local Similarity 37.8%; Pred. No. 1.2e-25;
Matches 155; Conservative 70; Mismatches 126; Indels 59; Gaps 12;

QY 756 EEETKNMGMREVLNLTKELELSKRYSRALRPSGNGRRMVDVPVASTGVQTEAVCGDAAE 815
Db 3 QQENNRDLGREIENLTKELELYRHFSKSLRPSLNGRRISDPQVFSKEVQTEAV-----D 57

QY 816 EETP-----AVFIRKSFOE-ENHIMSNLRQVGLKKPMERSSVL-----DRYPPAA 859
Db 58 NEPPDYKSLIPLERAVINGQLYEESNQD-----EDPNDEGSVLSFKCSQSTPCPV 108

QY 860 NELTMRKSWIPWMRKRE----NGPSTPQEKGRPRNQAGHPGELVLAPKQGQPLHIRVTP 915
Db 109 N-----RKLWIPWMKSKEGHLQNG-----KMQTKPNANFVQPGDLVLSHTPGQPLHIKVT 159

QY 916 DHENSTATLEITSPTSE--EFFSSTTVIPTLGNQKPRITIIIPSPNVMSQPK-SADPTLG 972
Db 160 DHVQNTATLEITSPTTSPHYSYTSYTAVIPNCGTPKQRTILQNASITPVKSKTSTEDLMN 219

QY 973 PERAMSPVTITISREKSPGGRSAFADRPASPIQIMTVSTSAAPTEIAVSPESQEVPMG 1032
Db 220 LEQGMSPITMATFARAQTPESCGSLTPERTMSPIQVLAVTGSASSPEQGRSPTEISAK 279

QY 1033 RTILKVTPEKQTVPAVRKYNSNANIITTEDNKKIHLGSGQFKRSPGPAAEVSPVITVR 1092
Db 280 HAIFRVSPDRQSSWQFORSNSNSSSVITTEDNKKIHLGSPYMQA---VASPVRPASPSA 336

QY 1093 PVNVTAKEVSTGTVLRSRPNHLSSRPGASKVTSTITITPVTTSSSTRGTQ 1142
Db 337 PLQDNRTQGLINGALNKT-----TNKVTSSITITPTATPLPRQSQ 376

RESULT 15
US-10-309-851-22
; Sequence 22, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
; CURRENT APPLICATION NUMBER: US/10/309,851
; CURRENT FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-851-22

Query Match      10.2%; Score 618.5; DB 14; Length 386;
Best Local Similarity 37.8%; Pred. No. 1.2e-25;
Matches 155; Conservative 70; Mismatches 126; Indels 59; Gaps 12;

QY 756 EEETKNMGMREVLNLTKELELSKRYSRALRPSGNGRRMVDVPVASTGVQTEAVCGDAAE 815
Db 3 QQENNRDLGREIENLTKELELYRHFSKSLRPSLNGRRISDPQVFSKEVQTEAV-----D 57

QY 816 EETP-----AVFIRKSFOE-ENHIMSNLRQVGLKKPMERSSVL-----DRYPPAA 859
Db 58 NEPPDYKSLIPLERAVINGQLYEESNQD-----EDPNDEGSVLSFKCSQSTPCPV 108

QY 860 NELTMRKSWIPWMRKRE----NGPSTPQEKGRPRNQAGHPGELVLAPKQGQPLHIRVTP 915
Db 109 N-----RKLWIPWMKSKEGHLQNG-----KMQTKPNANFVQPGDLVLSHTPGQPLHIKVT 159
```

```

QY 916 DHENSTATLEITSPTSE--EFFSSTTVIPTLGNQKPRITIIIPSPNVMSQPK-SADPTLG 972
Db 160 DHVQNTATLEITSPTTSPHYSYTSYTAVIPNCGTPKQRTILQNASITPVKSKTSTEDLMN 219

QY 973 PERAMSPVTITISREKSPGGRSAFADRPASPIQIMTVSTSAAPTEIAVSPESQEVPMG 1032
Db 220 LEQGMSPITMATFARAQTPESCGSLTPERTMSPIQVLAVTGSASSPEQGRSPTEISAK 279

QY 1033 RTILKVTPEKQTVPAVRKYNSNANIITTEDNKKIHLGSGQFKRSPGPAAEVSPVITVR 1092
Db 280 HAIFRVSPDRQSSWQFORSNSNSSSVITTEDNKKIHLGSPYMQA---VASPVRPASPSA 336

QY 1093 PVNVTAKEVSTGTVLRSRPNHLSSRPGASKVTSTITITPVTTSSSTRGTQ 1142
Db 337 PLQDNRTQGLINGALNKT-----TNKVTSSITITPTATPLPRQSQ 376
```

Search completed: September 7, 2004, 13:29:06
Job time : 162 secs

THIS PAGE BLANK (REF)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: September 7, 2004, 13:08:23 ; Search time 51 Seconds
(without alignments)
2285.963 Million cell updates/sec

Title: US-10-788-793-2
Perfect score: 6082
Sequence: 1 MRSRNQGGSSSNHGVSCP.....KIELKKSAAASSTASLGCGKG 1212

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	476	7.8	2094	2 S33124	tpr protein - huma
2	474.5	7.8	2139	2 T18296	myosin heavy chain
3	471.5	7.8	2007	1 B43402	myosin heavy chain
4	469.5	7.7	1961	1 A61231	myosin heavy chain
5	469	7.7	2663	1 S28261	centromere protein
6	464.5	7.6	1790	2 S67593	transport protein
7	463.5	7.6	2020	2 T21174	hypothetical prote
8	462.5	7.6	1909	2 A45592	liver stage antige
9	461.5	7.6	1999	1 S21801	myosin heavy chain
10	457	7.5	1164	2 T24806	hypothetical prote
11	457	7.5	1938	1 A40997	myosin heavy chain
12	450.5	7.4	1392	2 A43336	microtubule-vesicl
13	449	7.4	1938	2 JC5421	smooth muscle myos
14	449	7.4	1972	2 JC5420	smooth muscle myos
15	449	7.4	1992	2 A47297	myosin heavy chain
16	448.5	7.4	1964	2 A59282	nonmuscle myosin I
17	447.5	7.4	1959	1 A33977	myosin heavy chain
18	447	7.3	1427	2 S22695	restin - human
19	446.5	7.3	1039	2 S18199	myosin heavy chain
20	445	7.3	1972	1 A41604	myosin heavy chain
21	441	7.3	880	2 F75103	conserved hypothet
22	441	7.3	1979	1 S03166	myosin heavy chain
23	439	7.2	1690	2 T13030	microtubule bindin
24	439	7.2	1827	2 T16270	hypothetical prote
25	439	7.2	1940	1 A24922	myosin heavy chain
26	438	7.2	2346	2 T13829	Tpr homolog - frui
27	436.5	7.2	1407	1 S28589	trichohyalin - rab
28	435	7.2	1133	2 T22976	hypothetical prote
29	434	7.1	1976	2 A59252	myosin heavy chain

30	433.5	7.1	955	2 S24348	myosin heavy chain
31	433.5	7.1	1927	2 A59236	embryonic muscle m
32	433	7.1	1938	1 JX0178	myosin heavy chain
33	432.5	7.1	1597	2 S68420	citron - mouse
34	432.5	7.1	2253	2 T30336	nuclear/mitotic ap
35	432	7.1	1410	1 A57013	early endosome ant
36	431	7.1	2116	2 A26655	myosin heavy chain
37	430	7.1	1931	2 A59234	slow myosin heavy
38	430	7.1	1937	2 I38055	myosin heavy chain
39	429	7.1	1898	1 A45973	trichohyalin - hum
40	428	7.0	1940	1 S04090	myosin heavy chain
41	428	7.0	2017	1 A36014	myosin heavy chain
42	428	7.0	4687	1 A39638	plectin - rat
43	427	7.0	2057	2 S61477	myosin II heavy ch
44	426	7.0	1940	2 A29320	myosin heavy chain
45	425.5	7.0	1956	2 T16416	hypothetical prote

ALIGNMENTS

RESULT 1
S33124
tpr protein - human
N;Alternate names: kinase-related transforming protein (tpr-met); protein with promoter
C;Species: Homo sapiens (man)
C;Date: 22-Nov-1993 #sequence revision 26-May-1995 #text_change 11-Jan-2002
C;Accession: S33124; S23740; S00928; G01185
R;Mitchell, P.J.; Cooper, C.S.
Oncogene 7, 2329-2333, 1992
A;Title: The human tpr gene encodes a protein of 2094 amino acids that has extensive coi
A;Reference number: S33124; MUID:93064711; PMID:1437155
A;Accession: S33124
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-2094 <MIT>
A;Cross-references: EMBL:X66397; NID:G633225
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991
R;Mitchell, P.J.; Cooper, C.S.
Oncogene 7, 383-388, 1992
A;Title: Nucleotide sequence analysis of human tpr cDNA clones.
A;Reference number: S23740; MUID:92195670; PMID:1549355
A;Accession: S23740
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-725,'L' <MI2>
A;Cross-references: EMBL:X63105; NID:G37257; PIDN:CAA44819.1; PID:G37258
R;King, H.W.S.; Tempest, P.R.; Merrifield, K.R.; Rance, A.J.
Oncogene 2, 617-619, 1988
A;Title: Tpr homologues activate met and raf.
A;Reference number: S00928; MUID:88262257; PMID:3387099
A;Accession: S00928
A;Molecule type: mRNA
A;Residues: 1-31,'R',33-142 <KIN>
A;Cross-references: EMBL:Y00672; NID:G37255; PIDN:CAA68681.1; PID:G37256
R;Greco, A.
submitted to the EMBL Data Library, December 1995
A;Reference number: H00592
A;Accession: G01185
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 144-228 <GRE>
A;Cross-references: EMBL:X94208; NID:G1296797; PIDN:CAA63904.1; PID:G1296798
C;Genetics:
A;Gene: GDB:TPR
A;Cross-references: GDB:128821; OMIM:189940
A;Map position: 1q25-1q25
A;Introns: 177/3

Query Match 7.8%; Score 476; DB 2; Length 2094;
Best local Similarity 20.4%; Pred. No. 1.le-08;
Matches 314; Conservative 252; Mismatches 510; Indels 460; Gaps 67;

Db 1061 DEENKVKDLTNELOQTQLKLGETEKSALAAQVAATKKASDERDTLSQ-----NL 1108

QY 312 EHEEMNAKLANQESHNRQLRLKLVGLSQRIEEL-BTNK---SLQKAEELQELREKIAK 367

Db 1109 ENEKLTTK--NLTKADLEKKISGLKQDYEDLEDDKNKIEGDLRNAQRKIKELDDITK 1166

QY 368 GECGSSSLMAEVESLRKRVLEME-----GKDEEITKTEAQCRELKKLQJ---E 413

Db 1167 GADVSYLQKQKEEYESQIAKQEEKEAIGNDVKNKEKTIKEKELEIQSLQEKLDTEVE 1226

QY 414 EHSKELRLEVEKLQKRM-SELEKLEEAFFRSKSECTQLHLNLEKEKNLTDLLNELEV 472

Db 1227 KEDAEEKKKEIEKEMKALQEEKENVESSKNSTERDKKKLEDNLKDTQKKLDDMTADNEKL 1286

QY 473 KSRVKELECS-----ESRLEKAEEL-----SLK---DDLTKLKSFTVMLV 508

Db 1287 KAKAKDLEAQLNEVDNHEKAVADAELLNKKKAQSDKELNSLKAELEALTAKS-----V 1341

QY 509 DERKN-----MMEKIKQEEKVVDGLNKNFKVEQGKMDVTE-----KLIEE 549

Db 1342 VESKNKDSENEKALSEEDQANEKLNIOADLRKATADLQEAANEKKAEEVAQRDKLVAD 1401

QY 550 SKLLKLLKSEMEEKYSLTKERDELMGKLRSEERSCELSVDLLK-----KRL 599

Db 1402 NKQMTKTLEEIKARDEENTYKVENYKVLKRKEADLEAEANENLDIEKKDRMNKEQVKKL 1461

QY 600 DG-IEEVEREINRGSKGSEFTCPE--DNKIRELTLEIER---LKKRLQLEWVEGDL 652

Db 1462 EGELKETKDKLNAIAEKDSIFTAKKQSDADLELNKTVEEHDEVVAKLNTQITKLTRDN 1521

QY 653 MKTEDEYDQLEQFRTEQDKANFLSQLEEIHKQMAKHAIEKGEAV---SQEAELRHRF 709

Db 1522 QSAEELNELRSKADKDKKKISELEEQVNELESRPVGTGNADENEIKIRDAQIADLNKAL 1581

QY 710 RLEEAKSRDLQAEVQALKEKIHLMNK-----EDQLSQLQVDYSVLQORFMEEE 758

Db 1582 EMKGQVNNQLOATNKLKAKONDLTISKIEITENEMKKLENAKKRLEQDKDEADKAVSEQT 1641

QY 759 TKNNMGREVLNLTKELELSKRYSRALRPSG-----NGRRMVDVPVASTGVQTEAVCGDAA 814

Db 1642 IKRKGLEEVKKLTTEIQALKFQINA--PSSVAQEEKQORLESIDIAELKEQLEQERTTAA 1699

QY 815 EETPAVFIRKSFQEE-NHIMSNLRQV 840

Db 1700 NAEAE---RKIKQAEELDEVKFNLEDV 1722

RESULT 3

B43402

myosin heavy chain-B, neuronal - chicken

N;Contains: myosin ATPase (EC 3.6.4.1)

C;Species: Gallus gallus (chicken)

C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002

C;Accession: B43402; A43402

R;Takahashi, M.; Kawamoto, S.; Adelman, R.S.

J. Biol. Chem. 267, 17864-17871, 1992

A;Title: Evidence for inserted sequences in the head region of nonmuscle myosin specific yosin.

A;Reference number: A43402; MUID:92388144; PMID:1355479

A;Accession: B43402

A;Molecule type: mRNA

A;Residues: 1-2007 <TA>

A;Cross-references: GB:M93676; NID:g212448; PIDN:AAA48988.1; PID:g212452

A;Note: the sequence of residues 212-221 and 632-652 and the corresponding nucleotide se

A;Accession: A43402

A;Molecule type: mRNA

A;Residues: 1-211;222-631;653-2007 <TA2>

A;Cross-references: GB:M93676; NID:g212448; PIDN:AAA48985.1; PID:g212449

A;Note: sequence extracted from NCBI backbone (NCBIN:112864)

C;Comment: Alternatively spliced segments 1 and 2 are found exclusively in nonmuscle myo

C;Superfamily: myosin heavy chain; myosin motor domain homology

C;Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated

F;1-2007/Product: myosin heavy chain-B, neuronal #status predicted <MYN>

F;1-211;222-631;653-2007/Product: myosin heavy chain-B, nonmuscle #status predicted <MYS

F;88-802/Domain: myosin motor domain homology <MMOT>

F;178-185/Region: nucleotide-binding motif A (p-loop)

F;212-221/Region: alternatively spliced segment 1 #status experimental

F;559-593/Region: actin binding #status predicted

F;632-652/Region: alternatively spliced segment 2 #status experimental

F;692-714/Region: actin binding #status predicted

F;875-2007/Domain: coiled coil #status predicted <COI>

F;875-1315/Region: S2

F;1316-2007/Region: light meromyosin

F;129/Modified site: N6,N6-trimethyllysine (Lys) #status predicted

F;184/Binding site: ATP (Lys) #status predicted

F;732,742/Active site: Cys #status predicted

F;1954/Binding site: phosphate (Thr) (covalent) #status predicted

F;1987/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 7.8%; Score 471.5; DB 1; Length 2007;

Best Local Similarity 20.6%; Pred.No.1.4e-08;

Matches 259; Conservative 250; Mismatches 456; Indels 295; Gaps 49;

QY 67 KTKKSVLSKEDLIQLLSIMEGELQAREDVHMLRTEKTKPEVLEAHYCSAEPEKVL-- 124

Db 871 KVKPLQVTRQ-----EELQAKDEELMKVKEKQTKVE-----AELEEMERKH 913

QY 125 --VLHRDAILAQEKSIGEDVYEKPISELDRLEEKQKETYRRMLEQL-----LAEKCHR 176

Db 914 QQLLEKNILAEQLQAETELFAEAEMRRLAAKQE----LEILHDLESERVEEER 968

QY 177 RTVYELENEKHKHTDYMNKSDDFTNLLEOBERERLKKLLEQEKAYQARKEKENAKRLNKL 236

Db 969 NQI--LQNEKKKMQGHIQDLEE--QLDEEGARQKLQLEKVT-----EAKIKQME 1015

QY 237 DELVKLKSFALMLVDERQMHIEQLGQSKVQDLTQKLREEEKLVAVT----- 285

Db 1016 EEI-----LLEDQNSKFLKEKKLMEDRIAECTSQLAEEEKAKNLAKLNKQEMMIT 1068

QY 286 -----YKSKEDRQKLLK-----LEVDFEHKASRFQSQHEEMNAKLA 321

Db 1069 DLEERLKKEEKTQRQELEKAKRKLDGETTDLQDQIAELQAIEELKIQAKKEELQAALA 1128

QY 322 --NOESHNRQLRLKLV-----GLSQRIEELBETNKSLOKAEELQELREKIAKGECS 374

Db 1129 RGDEEAVQKNNAKLVIRELQAQIAELQEDLESEKASRNKAEKQKRDLSBEL-----EA 1181

QY 375 LMAEVESLRKRVLEMEGKDEEI-TKTEAQCRELKKLQEE-EHHSKELRLEVEKLQKRMS 432

Db 1182 LKTELED---TLDTTAAQQELRTKREQEVAELKKAIEEETKNHEAQIQ---EIRQRHAT 1234

QY 433 ELEKLEEAESRSKSECTQLHLNLEKEK-----NLTKDLLNELEVVK-----S 474

Db 1235 ALEELSEQLQAK---RFKANLEKNKQGLSDNKLACEVKVLQVKAESEHKRKKLDA 1290

QY 475 RVKELECSRSRLEKAELSLKDDLTCLKSFTVMLVDERKNMMEKIKQEBRKVDGLNKNFKV 534

Db 1291 QVQELTAKVTEGERLRLRVELAEKANKLQ-----NELDNVSSLLSEAEKKGKIPAKDAAS 1343

QY 535 EQGKVMVDVTEKLIIEESKLLKLS---EMEEKEYSLTKERDELMGKLRSEERSCELS 591

Db 1344 LESQLQDTQELLQEEETRQKLNLSRRIRQLBEEKNLQEQEEEEEAARKNLKQMLALQAO 1403

QY 592 VDLLKKRLD-----GIEEVEREINRGSKGSEFTCPEDNKIRELTLEIERLKKRLQ 644

Db 1404 LAEAKKKVDDDLGTIEGLEE-----NKKKLLKDMESL---SQRLSEKAMAYDKLETKNR 1455

QY 645 LEVVEGDLMKTEDEYDQLEQKFRTEQDKANFLSQLEEIHKQMAKH----- 690

Db 1456 LQOELDDLMDLDHQRQIVSNLEKKQKK---FDQMLAEKKNISARYAEERDRAEAAREK 1512

QY 691 --KAIEKGEAVSQEAELRHRF-----RLEEAKSRDLQAE 722

Db 1513 ETKALSLARALEEALEAKEEFERQNKQLRADMEDLMSSKDDVGKNVHLEKSK-RTLEQQ 1571

QY 723 VQALKEKIHLMNKK-----EDQLSQLQVDYSV-----LQQRFMEETKNKNMGREVLN 770

Db 1572 VEEMRTQLEBELEDELQATEDAKLRLEVNMQAMKAQAFERDLOARDEQNEEKKRMLVKQVRE 1631
QY 771 LTKELELSKRYSRALRPSGNGRRMVDVPVASTGVQTEAVCGDAEEETPAVFIR-KSFQE 829
Db 1632 LEAELE-DEKQRALAVAARKKMEMDLKLBEGQIEAANKARDEAIKQLRKLQAOQMDYOR 1690
QY 830 E-----NHIMSNLRQVGLK-KPMERSVLDTRYPPAANELTMRKSWIPWMRKRENGPS 880
Db 1691 ELEEARASRDEIFAQSESEKKLKGLE-----ABILQLQEEFAASERARRHAEQ 1739
QY 881 TPQEKGPFPNQAGHPGELVLAPKQGPLHVRVTPDHENSTATLEITTSPTSEFFSSTIV 940
Db 1740 ERDELADEIANSASGKSALLDEKRRLEARIAQLEEELEEEQSNMEL---LNERFRKTTIQ 1796
QY 941 IPTLGNOKPRITIIIPSPNVMSQPKSADPTLGPERAMSPVTITTISREKSPGEG-RSAPA 999
Db 1797 VDTLNSE-----LAGERSAAQKSENARQQL--ERQNKELK----AKLQELEGSVKSPF- 1843
QY 1000 DRPASPIQIMTVSTSAAPTEIAVSPESQEVPMGRITILKVTPPEK-QTVPAVRKYNSNANI 1058
Db 1844 ----KATISTLEAKIAQLEEQEAKERAANKLVRTTEKKLKEVFMQVEDERRHADQ 1898
QY 1059 ITTEDNKIHLGSGQKSPGPAAGVSPVITVRPNVNTAEKEVSTGTVLRSPRNHLSSR 1118
Db 1899 YKEQMEKANARM-KQLKRLQLEEAEEEAATRANASRR---KLQRELD DAT-----EAN 1945
QY 1119 PGASKVTSTITIT-----PVTTSSSTRGTQ-----SVSGQDGSSQRPTP 1156
Db 1946 EGLSRVSTLKNRLRRGGPITFSSSRSGRRQLHIGASLELSDDDDAESKGSVDVNEAQTTP 2005
RESULT 4
A61231
myosin heavy chain nonmuscle form A - human
N;Alternate names: cellular myosin heavy chain; myosin type 9; NMHC-A
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Homo sapiens (man)
C;Date: 12-May-1994 #sequence revision 14-Jul-1994 #text_change 19-Apr-2002
C;Accession: A61231; A34876; I52562; I61692
R;Simons, M.; Wang, M.; McBride, O.W.; Kawamoto, S.; Yamakawa, K.; Gdula, D.; Adelstein, C.
Circ. Res. 69, 530-539, 1991
A;Title: Human nonmuscle myosin heavy chains are encoded by two genes located on different chromosomes
A;Reference number: A61231; MUID:91316803; PMID:1860190
A;Accession: A61231
A;Molecule type: mRNA
A;Residues: 1-715 <SIM>
A;Cross-references: GB:M69180; NID:g189029; PIDN:AAA61765.1; PID:g189030
R;Saez, C.G.; Myers, J.C.; Shows, T.B.; Leinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 87, 1164-1168, 1990
A;Title: Human nonmuscle myosin heavy chain mRNA: generation of diversity through alternative splicing
A;Reference number: A34876; MUID:90138958; PMID:1967836
A;Accession: A34876
A;Molecule type: mRNA
A;Residues: 715-1961 <SAE>
A;Cross-references: GB:M31013; NID:g189035; PIDN:AAA36349.1; PID:g189036
R;Toothaker, L.E.; Gonzalez, D.A.; Tung, N.; Lemons, R.S.; Le Beau, M.M.; Arnaout, M.A.; Blood 78, 1826-1833, 1991
A;Title: Cellular myosin heavy chain in human leukocytes: isolation of 5' cDNA clones, characterization, and expression
A;Reference number: I52562; MUID:92003925; PMID:1912569
A;Accession: I52562
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-52, 'BAI', 56-659, 'T', 661-868, 'T', 870-930, 'C', 932-1239, 'KG', 1242-1337 <RES>
A;Cross-references: GB:M81105; NID:g188988; PIDN:AAA59888.1; PID:g553596
R;Bement, W.M.; Hasson, T.; Wirth, J.A.; Cheney, R.E.; Mooseker, M.S.
Proc. Natl. Acad. Sci. U.S.A. 91, 6549-6553, 1994
A;Title: Identification and overlapping expression of multiple unconventional myosin genes in Dictyostelium
A;Reference number: A55758; MUID:94294418; PMID:8022818
A;Accession: I61692
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 182-218 <BEM>

A;Cross-references: GB:L29141; NID:g457249; PIDN:AAA20904.1; PID:g531134
C;Genetics:
A;Gene: GDB:MYH9
A;Cross-references: GDB:120216; OMIM:160775
A;Map position: 22q12.3-22q13.1
C;Superfamily: myosin heavy chain; myosin motor domain homology <MMOT>
C;Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotide binding site
F;84-764/Domain: myosin motor domain homology <MMOT>
F;174-181/Region: nucleotide-binding motif A (P-loop)
F;552-565/Region: actin binding #status predicted
F;837-1277/Domain: actin binding #status predicted
F;837-1938/Domain: coiled coil #status predicted
F;125/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
F;180/Binding site: ATP (Lys) #status predicted
F;694/704/Active site: Cys #status predicted
Query Match 7.7%; Score 469.5; DB 1; Length 1961;
Best Local Similarity 22.9%; Pred. No. 1.6e-08;
Matches 228; Conservative 181; Mismatches 341; Indels 247; Gaps 40;
QY 67 KTKKSVLSKEDLIQLLSIMEGELQAREDEVHMLRTEKTPKPEVLEAHYGSAPKVLRLV 126
Db 833 KVKPLQLQVSRQ-----EEMMAKEEL-----VKVREKQLAAENRLMEMETLQS 878
QY 127 HRDAILAQEKSIGEDVYEKPISEL-DRLEEKQKETYRRMLEQLLLAEKCH-RRTVYELEN 184
Db 879 MAEKILQQLQAEQELCAEAEELRARLTAKQEQ-----LEEICHDLAEARVEEEE 928
QY 185 EXKHHTDYMNKSDFTNLLQEQRERLKKLLEQEKAYQARKEKENA---KRLNKLRLDELVK 241
Db 929 ERYQHLQAEKKK-----MQQNIQEELEEEESARQKLOLEKVTEAKLKKLEEQII 982
QY 242 LKSPALMLVDERQWHIEQLGLQSQKVDLTQKLRBEEKLKAVTYKSKEDRQKLLKLEVD 301
Db 983 LEDQNKCLAKEKK-----LLEDRIAETTNLTBEEKSKSLA-KLKNKHEAMI---TD 1031
QY 302 FEHKASRFSQHEEMNAKLANQESHNRQLRLKVLGSORIEELE-----ETNKSLOKABEE 357
Db 1032 LEERLRREEKQROEL-----EKTRKLEGSTDLSDQIAELQAIKQIAELKQIAELK 1084
QY 358 LQELREKIAKGECSNSSLMAEVESILRKRVLME-----GKDEEIT 397
Db 1085 LQAALARVEEEAAQNMALKKIRELESQISELQEDLESERASRNKAEKQKRDGEELEAL 1144
QY 398 KTE-----AQCRELK-----KKLQEEHHHSKELRLEVEKLQKRMSELEKLEE 439
Db 1145 KTELEDTLDTAAQQLRSKREQEVNLLKTTLEEAQTHEAQIQ-EMRQKHSQAVEELAE 1203
QY 440 AFSRSKSECTQLHLNLEKEK---NLTKDLLNLELV-----VKSRVKELEC 481
Db 1204 QLEQTK---RVKANLEKAKQTLNERNERLANEVKVLQGGRDSEHKRKKVQAQLQELQV 1259
QY 482 SESRLEKAELSLLKDDLTCLKSFTVMLVDERKNNMMEKIKQEERKVDGLNKNFKVEQGKVM 541
Db 1260 KFNEGERVRTTELADKVTKLQV-----ELDNVTGLLSQSDSKSKLTGKDFSALESQ 1312
QY 542 VTEKLIEESKLLKSKSEMEEKEYSLTKERDELGMKLRSEERS-----CELSCSVD 593
Db 1313 TQELLQEEENRQKLSLSTKLKQVE---DEKNSFREQLEEEEEEAKHNLEKQIATLHAQVA 1368
QY 594 LKKR-----LDGIEEVEREINR---GRSCKGSEFTCPEDNKNIRELTLEIERLKKRLQ 643
Db 1369 DMKKKMWDSVGCLETAEVVKRLQKDLGLSQRHEEKVAAAYD-----KLEKTKTRLQ 1420
QY 644 Q-LE--VVEGD-----LMKTEDEYDQ-----EQKFRTEQDKANFLSQLEEEKH 685
Db 1421 QELDDLLVDLHQRSACNLEKKQKKFDQLLAHEKTIISAKYAEERDRAEAAREKETKAL 1480
QY 686 QMAKHKAIEKGEAVSQEAELRH---RFRLE-----EAKSRDLQAEVQ 724

Db 1481 SLA--RALE--EAMEQKAELERLNKQFRTEMEDLMSSKDDVGKSVHELEKSKRALEQQVE 1536

QY 725 ALKEKIHEL MNK-----EDQLSOLQVDYSV-----LQORFMEETKNNMGREVNLNT 772

Db 1537 EMKTQLEEELEDELQATEDAKLRLEVLNLOAMKAQAFERDLQGRDEQSEKKQLVRQVREME 1596

QY 773 KELELSKRYSRALRPSNGRRMVDVPVASTGVQTEAVCGDAAEETPAVFIRKSPFQEE-- 830

Db 1597 AELE-DERKQORSMAVARKKLEMDLK-----DLEAHIDSANKNDEAIAKQLRKLQAQMK 1649

QY 831 -----NHIMSNLRQVGLK-KPMERSSVLDRYPPAANELTMRKSWIPWMRKRE 876

Db 1650 DCMRELDDTRASREEILAAQAKENKCLKSMEAEMIQLQEEELAAAEAKRQA----- 1700

QY 877 NGPSTPQEKGRPNQAGHPGELVLAPKQGQPLHIRV 913

Db 1701 -----QQRDELADEIANSSGKGALALEEKRRLEARI 1732

RESULT 5

S28261

centromere protein E - human

N;Alternate names: centromere 312K protein; kinesin-related protein CENP-E

C;Species: Homo sapiens (man)

C;Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001

C;Accession: S28261

R;Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.

Nature 359, 536-539, 1992

A;Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.

A;Reference number: S28261; MUID:93024922; PMID:1406971

A;Accession: S28261

A;Molecule type: mRNA

A;Residues: 1-2663 <YEN>

A;Cross-references: EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865

C;Genetics:

A;Gene: GDB:CENPE

A;Cross-references: GDB:361164; OMIM:117143

A;Map position: 4q24-4q25

C;Superfamily: centromere protein E; kinesin motor domain homology

C;Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop

F;7-335/Domain: kinesin motor domain homology <KWOT>

F;86-93/Region: nucleotide-binding motif A (P-loop)

F;486-2183/Domain: coiled coil #status predicted <COI>

F;92/Binding site: ATP (Lys) #status predicted

Query Match 7.7%; Score 469; DB 1; Length 2663;

Best Local Similarity 20.3%; Pred. No. 2.4e-08;

Matches 279; Conservative 267; Mismatches 455; Indels 376; Gaps 62;

QY 33 SEDAKNNKANRKEED-----VMASTIKRHLKPSGESEKTK-----KSVE 73

Db 1382 SQSQEQSLNMKXEDNETTKIVSEMEQFKPKDSALLRIEIMGLSKRLQESHDMKSA 1441

QY 74 LSKEDLIQLLSIMEGEL-QAREDVIHMLRTE-KTKPEVLEAHYGSAPKVLRLHRDAI 131

Db 1442 KEKDDLQRLQEVLOSQSDQLKENIKEIVAKHLETEELKVAHCCLKEQEETINELR---V 1498

QY 132 LAQEKSIGEDVYEKPISEL-DRLEEKQKQETYRRMLEQLLAE-----KCHR 176

Db 1499 NLSEKETEISTIQKLEAINDKLQNKIQEIYEKE-EQLNIQI SEVQENVNELKQFKEHR 1557

QY 177 R----TVYELENEKHKHTDYMNKSDFTNLLEQERERLKKLLEQEKAYQARKE--KENAK 230

Db 1558 KAKDSALQSIESMLETNRLQESQEEIQIMIKEBEMKRV---QEALQIERDQLKENTK 1614

QY 231 RL-NKLRDELVK---LKSFALMLVDERQMHIQL--GLQSQKVQ-----DLTQKL 274

Db 1615 EIVAKMKESQEKEYQFLKMTAVNETQEKMCIEIHLKEQFETQKLNLENIETENIRLTQIL 1674

QY 275 REEEKLVKAVTYKSKEDR---QKLLKLEVD-----FEHKASRFSQEHHEEMNAKLANQESH 326

Db 1675 HENLEEMRSVT-KERDDLSRVSVEETLKVERDQLKENLRETITRDLEKQEEELKIVHMHLEH 1733

QY 327 NRQLRLKLVGL-SORIEELEETNKSLOKAAEELQ-----ELREKIAKGECGNSSLMAEVES 381

Db 1734 QETID-XLRGIVSEKTEISNMQKOLEHSNDALKAQDLKIQEELRIAHMHLKEQETIDK 1792

QY 382 LRKRVLMEGK-----DEEITKTEAQCRELKKKLQEEHHSKELRLEVEKLQKRMSELEKL 437

Db 1793 LRGI VSEKTDKLSNMQKDLSENSAKLQEKIQELKANEHQLITLKQDVNETQKRVSEMEQL 1852

QY 438 -----EEAFSRKSECTQLHLNLEKEKNLTKDILNLELVVKS RVKE---LECSERLEKA 489

Db 1853 KKQIKDQSLTSLKLEIENL-----NLAQELHENLEEMKSVMKERDNLRRVBEETLKE 1904

QY 490 ELSLKDDLTKLKSFTVMLVDERKNMMEKIKQEERKVDGLNKNFKVEQGKVM DVTEKLI EE 549

Db 1905 RDQLKESLQETKARDLEIQQELKTARMLSKHEKKTVDKLRE-----KISEKTIQISDI 1957

QY 550 SKLLKLKSEMEEKEYSLTKERDELMGKLRSSEERSCELSVDLLKKRLDGI EEVEREI 609

Db 1958 QKDLDKSKDELQKKIQELQKKELQL---LRVKED-----VNMSHKKINEMEQ LKKQF 2006

QY 610 NRGSRCKGSEFTCPEDNKKIRELTLEIERLKKRLQOQLEVVEGDL MKTDEYDQLEQKFRTE 669

Db 2007 EPNYLCK-----CEMDN--FOLT--KKLHESLEEIRIV---AKERDELRRIRKESLKME 2052

QY 670 QDKANFLSQOLEEI-----KHQMAKHKAI-----EK-----GEAVSQEAELRH 707

Db 2053 RDQ--FIATLREMIARDRQNHQVKPEKRLSDGQOHLMESLREKCSRIKEL LKRYSEMDD 2110

QY 708 RFRLEEA KSRLQAEVQ-----ALKEKIHELMNKEDQLSOLQVDYSVLQ 751

Db 2111 HYECLNRLSLDLEKEIEFHRI MKLKYVLSYVTKIEEQHECINK-----FE 2157

QY 752 QRFMEETKNNMGREVNLTKLELSKRYSRALRPSGNGRRMVDVPVASTGVQTEAVCG 811

Db 2158 MDFIDEVEKQKELLIKIQHLQDQCDVPSREULDKLNQN---MDLHI-----EELK 2206

QY 812 DAAEETPAVFIRKSFQEEHNHMSNLRQV-----GLKKPMERSSVL 852

Db 2207 DFSESEFPS--IKTEFFQ---VLSNRKEMTQFLEEWLNTRFDIEKLKNGIQKENDRICQV 2261

QY 853 -----DRYP PAA NE-----LTMRKSWIPWMRKRENGPSTPQEKGRPNQAGHPGEL 899

Db 2262 NFFNRRIIAIMNESTEFERSATISKEW-----EQDLKSLKEKNEK----- 2303

QY 900 VLAPKQGOPLHIRVTPDHENSTATLEITSPTSEFFSSTTVIPTLGNQKPRITIIPSPNV 959

Db 2304 --LFKNYQTL-----KTSLASGAQVNPITQD-----NKNPHVT-SRATQL 2340

QY 960 MSQPKKSADPTL--GPERAMS-----PVTITTISREKSPGGRSAFADRPASP 1005

Db 2341 TTEKIRELENSLHEAKESAMHKEKSI IKMQKLEVTNDIIAKLQAKVHESNKCLEKTKET 2400

QY 1006 IQIMTVSTSAAPTEIAVSPESQEV-PMGRITILKVTPEKQTV PAVVRKYNSNANIITTEDN 1064

Db 2401 IQVLQDKVA-----LGAKPYKEEIEDLKMVLKIDLEKMKNAKEFEK-EISATKATVEYQ 2454

QY 1065 KIHHL-----GSQFKRSPGPAAGVSPVITVRPV-----NVTA 1098

Db 2455 KEVIRLLRENLRRSQQAQDTSVISEHTDPOPSNKP LTCGGSGIVQNTKALILKSEHRL 2514

QY 1099 EKEVS-----TGTVLRSPRNHLSSR-----PGASKVT 1125

Db 2515 EKEISKLKQQNEQLIKQKNELLSNNQHL SNEVKTWKERTLKREAHKQVTCENSPKSPKVT 2574

QY 1126 STIT----ITPVTTSSSTRGTQSVSGQDGSQRPTPTTRIP-----MSKGMKAGKPV 1171

Db 2575 GTASKKKQITPSQC-----KERNLQDVPVKESP KSCFFDSRSKSLPSPHPV 2620

RESULT 6

S67593

transport protein USO1 - yeast (Saccharomyces cerevisiae)

N;Alternate names: protein D2552; protein YDL058w

Db 986 LAKEKLEQQSD---LEQERLAKEKLGQQSDLEQER---LAKEKLGQQSDLEQERLA 1038
Qy 241 KLKSFALMLVDERQMHIEQLGLQSQKQVQDLTQKLREE---EKLKAVTYKSKEDR---QK 294
Db 1039 KEK-----LQEQSDLEQERLAKEKLEQQSDLEQERLAKEKLGQQSDLEQERLAKEK 1092
Qy 295 LLKLEVDFEHKASRFSQHEHEEMNAKLANQESHNRQLRLKLVGLSQRIEELEETNKSLOKA 354
Db 1093 LQGGQSDLEQER---RLAKE-----KLQGGQSDLEQERLAKEKLGQQSDLEQERLAKEKL 1144
Qy 355 BEELQEL-REKIAKGECGNSLMAEVESLRKRVLEMEGKDDEITKTE-----AQ 402
Db 1145 QEQSDLEQERLAKEKLEQQSDLEQERRAKEKLEQQSDLEQERTKASKETLQEQQSDLEQ 1204
Qy 403 CRELKKKLEQEEHHSKELRLEVEKLOKRMSEL-----EKLLEAFS-----RSKSECT 449
Db 1205 ERLAKEKLEQQSDLEQERRAKEKLEQQSDLEQERLAKEKLEQQSDLEQERRAKEKLEQ 1264
Qy 450 QLHLNLEKEKNLTKDL---LNELEVWKSrvKELECSSESRLKAEKLSLXDDTLTKLSFTVM 506
Db 1265 EQQSDLEQERRAKEKLEQQSDLEQERRAKEKLEQQSDLEQERLA-KEKLEQQS---- 1319
Qy 507 LVDERKNMMEKIKQBERKVDGLNKNFKVEQGVMDVTEKLEIESKLLK---LKSEMEEK 563
Db 1320 --DLEQERLAKEKLEQQSD-----LEQER--RAKEKLEQQSDLEQERLAKEKLEQ 1368
Qy 564 EYSLTKERDELMLGLRSEERSCELSVDLLKKRL--DGIEEVEREINRGRSCKGSEFT 621
Db 1369 QSDLEQER---RAKEKLEQQS-----DLEQDRLAKEKLEQQORDLEQERRAK----- 1413
Qy 622 CPEDNKIRELTLEIERLKKRLQQLVVEGDL-----MKTEDEYDQLEQKFRTEQDKAN 674
Db 1414 ----EKLQEQSDLEQERRAKEKLEQQSDLEQERRAKEKLEQQSDLEQERRA---KEK 1466
Qy 675 FLSQOLEETKHQMAKHAIEKGEAVSQEAEHLRFRLEAKSRDLQAEVQALKKEIHLM 734
Db 1467 LQEQSDLEQERLAKEKLEQQORDLEQER--RAKEKLEQQS--DLEQERRA-KEKLEQ-- 1520
Qy 735 NKEDQLSQLDVYSVL--QORFMEETKNMGMREVLNLTKELELSKRYSRALRPSGNG- 791
Db 1521 -QQSDLEQERLANEKLEQQORDLEQERRAKEKLEQ--QQSDLEQERRAKEKLEQQSDL 1576
Qy 792 ----RRMVDVPVASTGVQTEAVCGDAEETPAV-----FIRKSFQEEHNSNLRQVGLK 843
Db 1577 EQERRAKEKLEQQSDLEQERLAKEKLEQQORDLEQERLAKEKLEQQORDLEQ-RKADTK 1635
Qy 844 KPMER-----SSVL--DRYP-----PAANELTMRKSWIPWMRKRENGPSTPQEKGP 888
Db 1636 KNLERKKEHGDVLAEDLYGRLEIPAIELPSENE---RGYIIP-----HQSSILPQ 1681
Qy 889 PNQ 892
Db 1682 DNRG 1685

RESULT 9
S21801
myosin heavy chain, neuronal [similarity] - rat
N;Alternate names: myosin II
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 19-Apr-2002
C;Accession: S21801; PN0013; S18134
R;Sun, W.; Chantler, P.D.
J. Mol. Biol. 224, 1185-1193, 1992
A;Title: Cloning of the cDNA encoding a neuronal myosin heavy chain from mammalian brain
A;Reference number: S21801; MUID:92235856; PMID:1569576
A;Accession: S21801
A;Molecule type: mRNA
A;Residues: 1-1999 <SUN>
A;Cross-references: EMBL:X62659
R;Sun, W.; Chantler, P.D.

Biochem. Biophys. Res. Commun. 175, 244-249, 1991
A;Title: A unique cellular myosin II exhibiting differential expression in the cerebral c
A;Reference number: PN0013; MUID:91151356; PMID:1998509
A;Accession: PN0013
A;Molecule type: mRNA
A;Residues: 1914-1998, I' <SU2>
A;Experimental source: brain
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotide
F;84-763/Domain: myosin motor domain homology <MMOT>
F;174-181/Region: nucleotide-binding motif A (P-loop)
F;541-575/Region: actin binding #status predicted
F;653-675/Region: actin binding #status predicted
F;836-1999/Domain: coiled coil #status predicted <COI>
F;836-1276/Region: S2
F;1277-1999/Region: light meromyosin
F;125/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
F;180/Binding site: ATP (Lys) #status predicted
F;693,703/Active site: Cys #status predicted
F;1916/Binding site: phosphate (Ser) (covalent) #status predicted
F;1943/Binding site: phosphate (Ser) (covalent) #status predicted
Query Match 7.6%; Score 461.5; DB 1; Length 1999;
Best Local Similarity 24.3%; Pred. No. 3.1e-08;
Matches 210; Conservative 160; Mismatches 260; Indels 235; Gaps 40;
Qy 67 KTKSVLSKEDLIQLLSIMEGLOAREDVHMLRTEK-TKPEVLEAHYGSAPKVLK- 124
Db 832 KVKPLQLQVTRQ-----EEELQAKDEEL-MKKVVKQTKVE-----AELEEMERK 873
Qy 125 ---VLHRDAILAQEKSIGEDVYEKPISELDRLEEKQKETYRRMLEQLL-----LAEKCH 175
Db 874 HQLLLEKNILAEQLQAETELFAEAEEMRARLAQKQE-----LEEILHDLESRVEEEE 928
Qy 176 RRTVVELENEKHKHTDYMKNKSDDFTNLLEQERERLKKLLEQEKAYQARKEKENAKRLNKL 235
Db 929 RNQI--LQNEKKKEQGHKNDLEEQDEMESARQKL-----QLEKVTTEAK-LKKL 975
Qy 236 RDELVKLKS FALMLVDERQMHIEQLGLQSQKQVQDLTQKLREEEEEKLKAVTYKSKEDRQKL 295
Db 976 EEEQILEDNQNKLAKEKK-----LLEDRIAFTNLTTEEEESKSLA-KLKNKHEAM 1027
Qy 296 LKLEVDFEHKASRFSQHEHEEMNAKLANQESHNRQLRLKLVGLSQRIEELE-----ETNKS 351
Db 1028 I---TDLEERLRREKQREL-----EKTRRKLEGGDSTDLSQIAELQAIQIAELKMQ 1077
Qy 352 QKABEELQELREKIAKGECGNSLMAEVESLRKRVLEME-----G 391
Db 1078 AKKEEELQAALARVEEAAQKNMALKKIRELESQISELQEDLESERASRNKAQKRDIG 1137
Qy 392 KDEEITKTE-----AQRELK-----KKLQEEHHSKELRLEVEKLOKRMSE 433
Db 1138 EELEALKTELEDLTDTAAQQELRSKREQEVNLIKKTLEEEAKTHEAQIQ-EMRQKHSQA 1196
Qy 434 LEKLEEAFFSRKSSECTQLHLNLEKEK-----NLTKDLLNELEV-----VKSR 475
Db 1197 VEELAEQLEQTKRKVA---NLEKAKQTLNENGERBELANEVKVLQGGRDSEHKRKVYEAQ 1252
Qy 476 VKELECSERLEKAEKLSKDDLTKLKSFVTVMLVDERKNMMEKIKQBERKVDGLNKNPKVE 535
Db 1253 LQELQVKFNNEGERRVTELDKVTKLQV-----ELDNVTGLLSQSDSKSKLTQDFSA 1305
Qy 536 QGVMDVTEKLEIESKLLKLSKEMEKEYSLTKERDELMLGKLRSSEERS-----CE 587
Db 1306 ESQQLDTEQLQENRQKLSLTKLKQVE---DEKNSFREQLSEEEEAHKNLEKQIAT 1361
Qy 588 LSCSVDLLKKR-----LDGIEEVEREINR---GRSCKGSEFTCPEDNKIRELTLEIER 637
Db 1362 LHAQVADMKKKNMEDSVGCLTAEVVKRKLQKDLLEGLSQRHEEKVAAYD-----KLEK 1413
Qy 638 LKKRLQO-LE---VVEGD-----LMKTEDEYDQ-----EQKFRTEQDKAN 674
Db 1414 TKTRLQQLDLDLVDLDHQRQACNLEKKQKQKFDQLLAEIITSAKYAEERARDAERAE 1473

QY 675 FLSQQLSEIKHQMAKHAIEKGEAVSQEAE-LRHRF-----RLEEAKSR----- 717
Db 1474 KATKELSLARAEL-----EAMEQKAFLRKNLQEMTERLDEMSSKVDDAKSVLEHE 1524
QY 718 --DLQAEVQALKEKIHLMNKEDQLSQ-----LQVDYSVLQORFM-----EEETK 760
Db 1525 KSKLGRQOVMEETK-QLLEEDELAAQTEDAKLRLEVNLAQMAKQAFERDLQGRQDDSEK 1583
QY 761 NKNMGREVLNLTKELE-LSKRYSTRA 784
Db 1584 QKKLVROVREMAELEDDQKEMSTRA 1608

RESULT 10
T24806
hypothetical protein T10G3.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24806
R;Burton, J.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19937
A;Accession: T24806
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1164 <WIL>
A;Cross-references: EMBL:Z81118; PIDN:CAB03330.1; GSPDB:GN00023; CESP:T10G3.5
A;Experimental source: clone T10G3
C;Genetics:
A;Gene: CESP:T10G3.5
A;Map position: 5
A;Introns: 15/3; 73/3; 387/3; 412/3; 708/3; 818/3; 894/3; 982/3; 1080/2

Query Match 7.5%; Score 457; DB 2; Length 1164;
Best Local Similarity 23.5%; Pred. No. 2.4e-08;
Matches 222; Conservative 183; Mismatches 304; Indels 234; Gaps 44;

QY 26 SDGGKGPSEDAKQNKANRKEEDVMASGTIKRHLKPSG-----ESEKTKKSVELSKEDLIQ 81
Db 204 SESGKVMEDLK-----RELKVVKSDVVRYEIEVSRLEKMLDQRPSEDDVNVLRTELVN 257
QY 82 LLSIMEG-----ELQAREDVIHMLRTEKTKPEV----LEAHYGSAPPE-KVLRVLH-- 127
Db 258 AQKLMDAISQEKDIEIKHEHLSIRNLSMEREKQHIVNENLEKKIGEETVKQLQISYDA 317
QY 128 -----RDAILAQEKSIGEDVYEKPISE----LDRLEEKQKETY--RRML----- 165
Db 318 QSEELKQORNVVVQLEARIEENVFE--LSENKQNVKRLKEDKVQESQDALQMLSNINGSNE 375
QY 166 EQLL-LAEKCHRRT-----VYE----LENEKHK-----HTDYMKNKSDFTNLLEQER 207
Db 376 EQMISLNSKFERNTAERKRIEAVFEKVTVQGERLKTLEMANLDTNLEASMGSLDDKER 435
QY 208 -----ERLKKLLEQEKAYQARKEK--ENAKRLNKLRLDELV-----KLKSFAL 247
Db 436 SLLEEKNEISERDSSINDLKEKLAESEKATKYKNELKEHADLVENLTQLNKLQENSK 495
QY 248 MLVDE-----RQMHIEQL-----GLQSQKVQDLTQKLREEE 278
Db 496 DLMEKISAGEGGAKMAIEQLEQEKVKLTNELQTSSEKTKKASGELEAKISELEKLRDAE 555
QY 279 EKLKAVTYKSKEDRQ----KLLKLEVDFEHKASRFSQEHHEEMNAKLANQESHNRQLRLK- 333
Db 556 ASRTDKEQKWQKEKESFERKLAEADEIKRKGRRFVMEKEMEEE--RQKATDRTLKLKD 613
QY 334 -LVGLSQRIE----ELEETNKSLOKAAEEELQELREKIAKGECCGNSSLMAVESLRKRVLE 388
Db 614 ALVNSEKNLETIKKSESDREKIVREKDAHLEENKKRI-----EDAVQKLEEAKEKRARE 666
QY 389 ME----GKDEEITKTEAQCRELKKLQEEHHHSEKELRLEVEKLRKMS-----ELEKL--- 437
Db 667 LEASVSSRDTTIVSTKSELSSELKGLTESNSFTIEELKVQVEKVSNEISEKQOEVENLMAE 726

QY 438 -----EEAFSRKSSECTQLHLNLEKEKNLTKDLLNELEV 471
Db 727 MRDKEAHWKTKRDEFEAQMRLRNOEDNEEASSTLKSVEQEQ----LMKEKETSGEKNQLIS 782
QY 472 VKSRVKELECSSESRLKAELSLKDDLTCLKSFTVWMLVDERKNM---MEKIKQE-----ERK 524
Db 783 VKSOLEELKTEVERLIRSEEEKTQEIEKLSAVTATTQERDELATATSESLRTECENLNSK 842
QY 525 VDGLNKNFKVEQGVMDVTEKLIIEESKLLKLKSEMEEKEYSLTKERDELGMGLRSEER 584
Db 843 IQSIEESRRHABEKGSENLERMITKSRLEK---DIEERESTIQSIOEALETKDNEIESL 899
QY 585 SC-----ELSCSVDLLKKRLDGIIEEVEREINRGRSC-----KGSE-----FTCPE 624
Db 900 KTTQRVVEDELVSKISHIESFNSRIEEFEKEMASGKRKTIERLEAEKAEETKLVVFTGTQ 959
QY 625 DNKIRELTLEIERLKKRLQQLLEVVEGDLMKTEDEYDQLEQKFR-TEQDKANFLSQQLEEI 683
Db 960 SQQEEEL-----EKLQKEIQEKETT---IARMTSSKTQFEAMFADVQQTLSKEINDKTEEI 1012
QY 684 KHQMAKHKAIEKGEAVSQEAEELRHRFRLEEAKSRDLQAE--VQALKEKIHLMN-KEDQL 740
Db 1013 ERLMERIDSLEK-----VNHS-RIEELESRLTORERVVESLEADLAAVRNIEQEKL 1062
QY 741 SOLQVDYSVLQORFMEEBEETKNKMGREVNLNLTKELEL-SKRYSTRA 782
Db 1063 DELQ-----KLKEEF---DGKFKYFKSRILENSKKLQIDADRYSTRA 1098

RESULT 11
A40997
myosin heavy chain, striated adductor muscle - scallop (Aequipecten irradians)
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Aequipecten irradians
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C;Accession: A40997; S13557
R;Nyitrai, L.; Goodwin, E.B.; Szent-Gyorgyi, A.G.
J. Biol. Chem. 266, 18469-18476, 1991
A;Title: Complete primary structure of a scallop striated muscle myosin heavy chain. Seq
A;Reference number: A40997; MUID:92011595; PMID:1917970
A;Accession: A40997
A;Molecule type: mRNA
A;Residues: 1-1938 <NYI>
A;Cross-references: GB:X55714; NID:G5611; PIDN:CAA39247.1; PID:G5612
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; muscle contraction; nucleotide b
F;86-763/Domain: myosin motor domain homology <MMOT>
F;176-183/Region: nucleotide-binding motif A (P-loop)
F;547-586/Region: actin binding #status predicted
F;653-675/Region: actin binding #status predicted
F;836-1938/Domain: coiled coil #status predicted <COI>
F;836-1276/Region: S2
F;1277-1938/Region: light meromyosin
F;182/Binding site: ATP (lys) #status predicted
F;693,703/Active site: Cys #status predicted

Query Match 7.5%; Score 457; DB 1; Length 1938;
Best Local Similarity 22.6%; Pred. No. 4.2e-08;
Matches 224; Conservative 205; Mismatches 382; Indels 178; Gaps 41;

QY 64 SEKTKKSVELSKEDLIQLLSIMEGELQAREDVIHMLRTEKTKPEVLEAHYGSAPPEKVL 123
Db 831 SKVKPLLSARQEEEMKEQLKQMD---KMKED---LAKTERIKKEL-----EEQNVT 876
QY 124 RVLHRDAILAQ-----EKSIG--EDVYEKPISELDRLEEKQKETYRRMLEQ----- 167
Db 877 LLEQKNDLFLQLQTLEDSMGDQEEERVEKLIQMKADEFESQIKELEERLLDEEDAAADLEGI 936
QY 168 --LLLAEEKCH-RRTVYLENEKHK-HTDYMKNKSDFTNL---LEQERERLKKLLEQEKAY 220
Db 937 KKKMEADNANLKKDIDGDLTQLQAEQDKAHKDNQISTLQGEISQQDEHIGKLNKEKKAL 996

[illegible]

Dd	1177	--DADEEKASLQKSISITSALLTEKDAELEKLNEVTVLRGENASAKSLHSVVQTLES DK	1234
QY	695	KGEAVSGEAE LHRFRLEEA KSRDLQAEVQA LKEKH IHELMNKEDQLSQ QVDYSVLQQR F	754
Dd	1235	VKLELVKNL-----ELQ LKENKRQLSSSCNTDTQAD	1267
QY	755	MEETKQNMGR EVNLTKLELSKRYSRALRPSGNGRRMVD	796

Query Match				7.4%;	Score 450.5;	DB 2;	Length 1392;
Best Local Similarity				22.9%;	Pred. NO. 4.8e-08;		
Matches 202;				Conservative 183;	Mismatches 316;	Indels 181;	Gaps 38;
QY	30	KGPSEDA---KKNKANRKEED----	VMASGTIKRHL---KPSGESEKTKKSVELSKEDL	79			
Db	452	KGDLEAVTSEKSRIMELEKDIALRVQEV	AEALRRRLESNKPAGD-----VMS	499			
QY	80	IQLLSIMEGELQAREDVIIHMLRTEKTKPEV-	LEAHYGSARP--EKVLRVILHRDAILAQE	135			
Db	500	---LSSLQ-EISSQLEKLEVRTTDHQR-EITSLKEHFGAREETHQKEIKALY----	TATE	550			
QY	136	KSIGEDVVEKPISELDRLEBEKQKETYRRML	EQLLLAECKHRRTVYELENEBKHK-----	188			
Db	551	KLSEN--ESLSKLEHANKENSVDIALWKS	KLETAIASHQQAAMEELKVSFSKGLGTETA	608			
QY	189	-----HTDYMNSKDDFTNLLEQERERL	KKLLE--QEKAYQARKEKENA-----	229			
Db	609	EFAELKTQIEKMRLDYQHEIENLONQO	SERAAHAKEMEAALRAKLMKVIKEKENSLEAIR	668			
QY	230	KRLNKLRE-LVKLKSFALMLVDERQMHTIEQ	LGL-----OSQKVQDLTQKLRREEEK-	280			
Db	669	SKLDAEQDHLVEMED-TLNKLQEA	EAIKVKELEVLQAKCNEQTVIDNFTSOLKATEEKL	727			
QY	281	--LKAVTYKSKEDRQKLLKLEVD	FEHKASRFSQEHEEMNAKLANQESHNRQLR--	335			
Db	728	LDLDALRKASSEKSEMKKLRQQLEAAEK	QIKHLEIEKNAESSKASSITRELOQRELKLT	787			
QY	336	GLSORIEELBEETNSLOKABEELQELRE	KIANG-----ECGN-----SS	374			
Db	788	NLOENLSVSQVKETLEK--ELQILKEK	FAEASEEAVSVQRSMQETVNLKHQKEEQFNM	844			
QY	375	LMAEVESLRKRVLEMEGKDDEEITKTEA	CQRELKKLQEE-----	413			
Db	845	LSSDLEKLRENLADEAKFREKDEREEQ	LIKAKEKLENDIAEIMKMSGDNSSQLTKWNDE	904			
QY	414	----EHSKELRLEVEKLQKRMSLEKLE	EAFS----RSKSECTQLHLNLEKEKNLT	465			
Db	905	LRLKERDVEELQLKLTKANENASFLQ	SIEDMTVKAEQSQEAAKKH---EEEKELERK	961			
QY	466	LNELEVYKSRVKELECSERLE--KABLS	LKDDLTKLKSFVTVMLVDERKQNMMEKIKOBER	523			
Db	962	LSDLE-----KMETSHNQOELKARYERAT	SETKTKEEIL-----QNLQKTLLDTE	1010			
QY	524	KVDGLKNFKVEQGVMDVTE--KLJBES	KLKLKXSEMEEKEYSLTKERDELMGKL	581			
Db	1011	KLKGARBE--NSGLLQLEELRQKAEKAKA	QAQTAEADAMQIME-QMTKEKTETLASL	1066			
QY	582	EERSCELSCSVDLLKKRLDGI	EEVEREINRGRSCKGSEFTCPEDNKIRELTLEIERLK--	639			
Db	1067	KQTNAKLQNLQELDTLKE--NNLKNVE	-ELN-----KSKELLTVENQKMEEF	1118			
QY	640	--KRLQOLEVVEGDLMKTEDEYDQLE	QKFRTEQ---DKANFLSQOLEEIKHQMAKHKAIE	694			
Db	1119	AAQKSQLSALQOEENVKLA	EELGRSRDEVTSHQKLEBERSVLNNQNLLEMKKRESKF	1176			
QY	695	KGEAVSQEAELRHRFRLEEAKSRDLQAE	VOALKEKITHELMNKEDQLSQLQVDYSVLQORF	754			
Db	1177	--DADEKASLQKSITSALLTEKD	AELEKLRNEVTVLRGENASAKSLHSVQVLTLESDK	1234			
QY	755	MEEETKNKNMGREVLNLTKELELSK	RYSRALRPSGNRRMVD	796			
Db	1235	VKLELKVKNL-----ELQLENKRQ	LSSSSGNTDTQAD	1267			

RESULT 13
JC5421
smooth muscle myosin heavy chain 2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 02-Feb-2001
C;Accession: JC5421
R;Hasegawa, K.; Arakawa, E.; Oda, S.; Matsuda, Y.
Biochem. Biophys. Res. Commun. 232, 313-316. 1997

A;Title: Molecular cloning and expression of murine smooth muscle myosin heavy chains.
A;Reference number: JC5420; MUID:97242182; PMID:9125171
A;Accession: JC5421
A;Molecule type: mRNA
A;Residues: 1-1938 <HAS>
A;Cross-references: DDBJ:D85924; NID:g1945079; PIDN:BAA19691.1; PID:g1945080
A;Experimental source: smooth muscle
C;Comment: This protein plays a role in smooth muscle cell contraction.
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: nucleotide binding; P-loop
F;88-771/Domain: myosin motor domain homology <MMOT>
F;178-185/Region: nucleotide-binding motif A (P-loop)

Query Match 7.4%; Score 449; DB 2; Length 1938;
Best Local Similarity 23.5%; Pred. No. 7.8e-08;
Matches 203; Conservative 167; Mismatches 292; Indels 202; Gaps 33;

QY	68	TKKSVELSKEDLIQ-----LLSI--MEGELQAREDVIHML--	100
Db	807	TKRQQQLTAMKVIQRNCAAYLKLNRWQWRLFTKVKPLLVQVTRQEEEMQAKEEMQKITE	866
QY	101	RTEKTKPEVLEAHYGSAPPEKVLRLVLRDAILAQEKSIGEDVYEKPISELDRLEEKQKET	160
Db	867	RQKAETELKELE-----QKHTQLAEKTL---LQEQLOAETELYAEESEMR	910
QY	161	YRRMLEQLLLAEKCHRRTVYELENEKHKHTDYMNSDDFTNLLEQERER-----LKKL	213
Db	911	VR-----LAACK-----QELEEILHEMEARLEEEEDRRQQLQAEKVKMAQQMLDLEE	958
QY	214	LEQEKAYQARKEKENA---KRLNKLRLDELVKLSFALMLVDERQMHIQGLQSQKVQDL	270
Db	959	LEEEAARQKLQLEKVTAEAKIKKLEDDI-----LVMDQNSKLSKERKLLLEERVSDL	1011
QY	271	TQKLREEEEEKLKAVT-----YKSKEDRQKLLKLEVDFEHKASRFSQEH	313
Db	1012	TTNLAAEEEEKAKNLTKLKSKHESMISELEVLKKBESRQLEKLRKLEGDASDFHEQI	1071
QY	314	EEMNAKLANQESHNRQLRLKL-----VGLSQRIEELEETNKSLOK-----AE	355
Db	1072	ADLQAOIA-----ELKMQLAKKEEELQAAALARLDEEIAQKNALKKIRELEGHISDLQ	1124
QY	356	EELQELREKIAKGECCGSSSLMAEVESLRKRV---LEMEGKDEEI-TKTEAQCRELKKKLQ	411
Db	1125	EDLDSERAARNKAQKRDLGEELEALKTELEDLTSTATQOELRAKREQEVTVLKKALD	1184
QY	412	EE-EHHSKELRLEVEKLRMSSELEKLEEAFAFSKSECTQLHLNLEKEKNLTKDOLLNELE	470
Db	1185	EETRSHEAQVQEMRQKHTQAVEELTEQLEQFKRAKANLDKSKQTLEKE---NADLAGELR	1241
QY	471	V-----VKSrvKLE-----CSESRLKAEKSLKDDLTCLKSFTVMLVDERKNM	514
Db	1242	VLGQAKQVEVHKKKLEVLQDLQSKSCDGERARAELS--DKVHKLQ-----NEVESV	1292
QY	515	MEKIQEERKVDGLNKNFKVEQGKVMVTEKLIIESKKLLKLS---EMEEKYSLTKER	571
Db	1293	TGMLNEAEGKAIKLAKDVASLSGSLQDTQELLQETRQKLNVTKLRLQLEDERNSLDQDL	1352
QY	572	DELMGKLRSSEERSCELSVDLLKKRL-----DGIEEVEREINRGRSCKGSEF	620
Db	1353	DEEMEAQNLERHVSFTLNQLSDSKKKLQDFASTIEVMEEGKKRLQKEM-EGLSQOYE	1411
QY	621	TCPED-----NKI-----RELTLIEIRLKKRLQQLLEVVEGDL-MKTEDEY	659
Db	1412	AAAYDKLETKNRLQQLDLDLVLDNRQLVSNLEKKQKQFDQLLAEKNISSKYADER	1471
QY	660	DQLEQKFRTEQDKANFLSQOLE---EIKHQMAKHKAIEKGE---AVSQEALRHRFRLEE	713
Db	1472	DRAEAAREKETKALSLARALEEALEAKEELERTNKMKAEMEDLVSSKDDVGNVHVELE	1531
QY	714	AKSRDLQAEVQALKEKIHE---LMNKEDQLSLOQVDYSV-----LQORFMEETKN	761
Db	1532	KSRALETQMEEMKTQLEESDDVQATEDAKLRLVNMQALKGQFERDLQARDEQNEEK	1591

QY	762	KNMGREVLNLTKELELSKRYSRAL	785
Db	1592	RQLQRQLHEYTELE-DEKQORAL	1614

RESULT 14
JC5420
smooth muscle myosin heavy chain 1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 02-Feb-2001
C;Accession: JC5420
R;Hasegawa, K.; Arakawa, E.; Oda, S.; Matsuda, Y.
Biochem. Biophys. Res. Commun. 232, 313-316, 1997
A;Title: Molecular cloning and expression of murine smooth muscle myosin heavy chains.
A;Reference number: JC5420; MUID:97242182; PMID:9125171
A;Accession: JC5420
A;Molecule type: mRNA
A;Residues: 1-1972 <HAS>
A;Cross-references: DDBJ:D85923; NID:g1945077; PIDN:BAA19690.1; PID:g1945078
A;Experimental source: smooth muscle
C;Comment: This protein plays a role in smooth muscle cell contraction.
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: nucleotide binding; P-loop
F;88-771/Domain: myosin motor domain homology <MMOT>
F;178-185/Region: nucleotide-binding motif A (P-loop)

Query Match 7.4%; Score 449; DB 2; Length 1972;
Best Local Similarity 23.5%; Pred. No. 7.9e-08;
Matches 203; Conservative 167; Mismatches 292; Indels 202; Gaps 33;

QY	68	TKKSVELSKEDLIQ-----LLSI--MEGELQAREDVIHML--	100
Db	807	TKRQQQLTAMKVIQRNCAAYLKLNRWQWRLFTKVKPLLVQVTRQEEEMQAKEEMQKITE	866
QY	101	RTEKTKPEVLEAHYGSAPPEKVLRLVLRDAILAQEKSIGEDVYEKPISELDRLEEKQKET	160
Db	867	RQKAETELKELE-----QKHTQLAEKTL---LQEQLOAETELYAEESEMR	910
QY	161	YRRMLEQLLLAEKCHRRTVYELENEKHKHTDYMNSDDFTNLLEQERER-----LKKL	213
Db	911	VR-----LAACK-----QELEEILHEMEARLEEEEDRRQQLQAEKVKMAQQMLDLEE	958
QY	214	LEQEKAYQARKEKENA---KRLNKLRLDELVKLSFALMLVDERQMHIQGLQSQKVQDL	270
Db	959	LEEEAARQKLQLEKVTAEAKIKKLEDDI-----LVMDQNSKLSKERKLLLEERVSDL	1011
QY	271	TQKLREEEEEKLKAVT-----YKSKEDRQKLLKLEVDFEHKASRFSQEH	313
Db	1012	TTNLAAEEEEKAKNLTKLKSKHESMISELEVLKKBESRQLEKLRKLEGDASDFHEQI	1071
QY	314	EEMNAKLANQESHNRQLRLKL-----VGLSQRIEELEETNKSLOK-----AE	355
Db	1072	ADLQAOIA-----ELKMQLAKKEEELQAAALARLDEEIAQKNALKKIRELEGHISDLQ	1124
QY	356	EELQELREKIAKGECCGSSSLMAEVESLRKRV---LEMEGKDEEI-TKTEAQCRELKKKLQ	411
Db	1125	EDLDSERAARNKAQKRDLGEELEALKTELEDLTSTATQOELRAKREQEVTVLKKALD	1184
QY	412	EE-EHHSKELRLEVEKLRMSSELEKLEEAFAFSKSECTQLHLNLEKEKNLTKDOLLNELE	470
Db	1185	EETRSHEAQVQEMRQKHTQAVEELTEQLEQFKRAKANLDKSKQTLEKE---NADLAGELR	1241
QY	471	V-----VKSrvKLE-----CSESRLKAEKSLKDDLTCLKSFTVMLVDERKNM	514
Db	1242	VLGQAKQVEVHKKKLEVLQDLQSKSCDGERARAELS--DKVHKLQ-----NEVESV	1292
QY	515	MEKIQEERKVDGLNKNFKVEQGKVMVTEKLIIESKKLLKLS---EMEEKYSLTKER	571
Db	1293	TGMLNEAEGKAIKLAKDVASLSGSLQDTQELLQETRQKLNVTKLRLQLEDERNSLDQDL	1352
QY	572	DELMGKLRSSEERSCELSVDLLKKRL-----DGIEEVEREINRGRSCKGSEF	620
Db	1353	DEEMEAQNLERHVSFTLNQLSDSKKKLQDFASTIEVMEEGKKRLQKEM-EGLSQOYE	1411

QY 621 TCPED-----NKI-----RELTLIERLKKRLOQLEVVEGDL-MKTEDEY 659
Db 1412 AAAYDKLEKTKNRLQQLDLDLVVDLNDQRLVNLKQKKFQDLAAEKNISSKYADER 1471
QY 660 DQLEQKFRTEQDKANFLSQOLE---EIKHOMAKHKAIEKGE---AVSQEAELRHRFRLEE 713
Db 1472 DRAEAAREKETKALSARALEFALEAKEELRTNMKLKAEMEDLVSSKDDVGNVHELE 1531
QY 714 AKSRDLQAEVQALKEKIHE----LMNKEDQLSQLOVDYSV-----LQRFMEETKN 761
Db 1532 KSKRALETQMEEMKTQLESEDDVQATEDAKLRLEVNMQALKGQFERDLQARDEQNEEKR 1591
QY 762 KNMGREVLNLTKELELSKRYSRAL 785
Db 1592 RQLQRQLHEYETELE-DEKQORAL 1614
RESULT 15
A47297
myosin heavy chain form B, nonmuscle - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 02-Feb-2001
C:Accession: A47297; A55441
R;Bhatia-Dey, N.; Adelstein, R.S.; Dawid, I.B.
Proc. Natl. Acad. Sci. U.S.A. 90, 2856-2859, 1993
A:Title: Cloning of the cDNA encoding a myosin heavy chain B isoform of Xenopus nonmuscul
A:Reference number: A47297; MUID:93219383; PMID:8464900
A:Accession: A47297
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1992 <BHA>
A:Cross-references: GB:I09740; NID:g214623; PIDN:AAA49915.1; PID:g214624
A:Experimental source: XTC cells
A>Note: sequence extracted from NCBi backbone (NCBIP:128722)
R;Kelley, C.A.; Oberman, F.; Yisraeli, J.K.; Adelstein, R.S.
J. Biol. Chem. 270, 1395-1401, 1995
A:Title: A Xenopus nonmuscle myosin heavy chain isoform is phosphorylated by cyclin-p34
A:Reference number: A55441; MUID:95138137; PMID:7836406
A:Accession: A55441
A:Status: preliminary
A:Molecule type: protein
A:Residues: 198-232 <KEL>
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: nucleotide binding; P-loop
F:88-787/Domain: myosin motor domain homology <MMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)
Query Match 7.4%; Score 449; DB 2; Length 1992;
Best Local Similarity 23.1%; Pred. No. 8e-08;
Matches 198; Conservative 181; Mismatches 329; Indels 150; Gaps 32;
QY 40 KANRKEEDVMASGTIKRHLKPSGESEKTKKSVEL-----SKEDLIQLLSIMEGELQARE 94
Db 862 QVTRQEEELVAKDEELLKVK-----EKQSKVEGELVDMEQKHQQLVERKNILAEQLHAET 916
QY 95 DVIHMLRTEKTKPEVLEAHYG--SAEPKVLRLVLRDAILAQEKSIGEDVYEKPISELDR 152
Db 917 ELF-----AEAEEMRAPLAIKQEMEELRDLE----IRMEEEEEERNQVLQN----- 959
QY 153 LEEKQKETRYRMLEQLLLAEKCHRRRTVYVELENEKHKHTDYMNKSDDFTNLLEQERERL-- 210
Db 960 -EKKKMQTHVQDLEEQDDEEAAQK----LQLEKVTAEAKIKKMEEDILVLEDQNSKFLK 1014
QY 211 -KKLEQ---EKAYQARKEKENAKRLNKLRLDELVKLSFALMLVDEROMHIEQLGLQSQK 266
Db 1015 EKKLLEERIAESTSQLAEEEEKAKNLAKLK-----NKQEMMISDLEERLKK 1060
QY 267 VQDLTQKLREEEEKLVKAVTYKSKEDRQKLLKLEVDPEHKASRFSQHEEMNAKLA--NQE 324
Db 1061 EEKTRQELEKAKRKLGET---TDFQDQIAELQAIIEELKLQAKKEEELQAAALARGDEE 1117
QY 325 SHNRQLRLKLV-----GLSQRIEELEBETNKSQKABEELQELREKIAKGECSNLSMAEV 379

Db 1118 VLQKNNTLKLVRLEQAQIAELQEDLESEKASRNKAQKRDLSSEL-----EALKTEL 1170
QY 380 ESLRKRVLEMEGKDEEI-TKTEAQCRELKKKLOEE-EHHSKELRLEVEKLQKRMSELEKL 437
Db 1171 ED---TLDTTAAQQELRTKREQVEAELRKSIIEETRHNHEAQIQ---EMRQROATALEEL 1223
QY 438 EEAFFSRKSECTQLHLNLEKEKNL----TKDLLNELEVVKSRVKELECSERLE--KAEL 491
Db 1224 SEQLEQAK----RFKVNLEKNKQSLESNDKELATEVKSQQMKAESEYKRRKKLEGVQVEL 1279
QY 492 -----SLKODLTKLKSFVTMLVDERKNMMEKI KQBERKVDGLNKNFKVEQGKVM DVT 543
Db 1280 HAKVLEGDRLRADMEKSS--KLQNELENVSSLLEEAEBKKGIKAKDVASMESQLQDTQ 1336
QY 544 EKLIEESKKLLKLKS---EMEEKEYSLTKERDELGMGLRSEERSCELSVDDLKKRLD 600
Db 1337 ELLQEETRQKLNQSSRIRQLEEEKNLQEQOEEEEEAARKSLEKQILSLQSLIEAKKKVD 1396
QY 601 -----GIEEVEREINRGSRCKSGSEFTCPEDNKIRELTLEIERLKKRLQQLEVEVEGDL 653
Db 1397 DEVGTTIEGLEEVKKLLKDTTEGLG-----QRLEEKIIAYEKLKTKNRLQQELDDLM 1448
QY 654 KTEDEYDQLEQKFRTEQDKANFLSQOLEEIKHOMAKH-----KAIEKGE 697
Db 1449 VDLDHQRQIVSNLEKKQKK---FDQLLAEEKNISARHAEERDRAEADAREKETKALSAR 1505
QY 698 AVSQEAELRHRF-RLEE--AKSRDLQAEVQALKEKITHELMNKEDQLSQLOVDYSVLQQR 753
Db 1506 ALDEALEAQDEFERLNLKQLRAEMEDLMSSKDDVGKNVHEL-----EKSKRALDQQVEEMR 1560
QY 754 FMEEETKNKNMGREVLNLTKELE---LSKRYSRALRPPSGNGRRMVDVPVASTGVQTEAVC 810
Db 1561 TQLEDEDELOGTEDAKLRLEVNMQAMKAQFERDLQTRDEQNEEKKRALVKQVRELEAEL 1620
QY 811 GDAABEETPAVFIRKSFQ 828
Db 1621 EDERQQRAMAVAIIKKLE 1638

Search completed: September 7, 2004, 13:17:32
Job time : 62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2004, 13:00:13 ; Search time 29 Seconds
(without alignments)
2176.173 Million cell updates/sec

Title: US-10-788-793-2
Perfect score: 6082
Sequence: 1 MRSRNQGGSSNGHVSCP.....KIELKKSASSTASLGCGKG 1212

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	476	7.8	2349	1 TPR_HUMAN	P12270 homo sapien
2	469	7.7	1960	1 MYH9_HUMAN	P35579 homo sapien
3	469	7.7	2663	1 CENE_HUMAN	Q02224 homo sapien
4	464.5	7.6	1790	1 US01_YEAST	P25386 saccharomyc
5	458.5	7.5	1961	1 MYH9_RAT	Q62812 rattus norv
6	457	7.5	1938	1 MYS_AEQIR	P24733 aequipecten
7	452	7.4	1972	1 MYHB_HUMAN	P35749 homo sapien
8	449	7.4	1972	1 MYHB_MOUSE	Q08638 mus musculu
9	447.5	7.4	1959	1 MYH9_CHICK	P14105 gallus gall
10	447	7.3	1427	1 REST_HUMAN	P30622 homo sapien
11	445	7.3	1972	1 MYHB_RABIT	P35748 oryctolagus
12	441	7.3	880	1 RA50_PYRAB	Q9uzc8 pyrococcus
13	441	7.3	1978	1 MYHB_CHICK	P10587 gallus gall
14	440	7.2	1411	1 EEAI_HUMAN	Q15075 homo sapien
15	439	7.2	1940	1 MYH3_RAT	P12847 rattus norv
16	436.5	7.2	1407	1 TRHY_RABIT	P37709 oryctolagus
17	436	7.2	1976	1 MYHA_BOVIN	Q27991 bos taurus
18	436	7.2	3210	1 CENF_HUMAN	P49454 homo sapien
19	435.5	7.2	1433	1 REST_CHICK	O42184 gallus gall
20	435.5	7.2	1976	1 MYHA_RAT	Q9jlt0 rattus norv
21	434	7.1	1976	1 MYHA_HUMAN	P35580 homo sapien
22	433.5	7.1	1102	1 MYSC_CHICK	P29616 gallus gall
23	433.5	7.1	1690	1 C190_DROME	Q9vjes drosophila
24	433	7.1	1938	1 MYSS_CHICK	P13538 gallus gall
25	432.5	7.1	1597	1 CTRO_MOUSE	P49025 mus musculu
26	431	7.1	1116	1 MYS2_DICDI	Q08799 dictyosteli
27	430.5	7.1	2238	1 GOA4_MOUSE	P91vw5 mus musculu
28	430	7.1	1937	1 MYH8_HUMAN	P13535 homo sapien
29	429.	7.1	1898	1 TRHY_HUMAN	Q07283 homo sapien
30	428.5	7.0	4473	1 PLE1_CRIGR	Q9ji55 cricetulus
31	428	7.0	1940	1 MYH3_HUMAN	P11055 homo sapien
32	428	7.0	2017	1 MYSN_DROME	Q99323 drosophila
33	428	7.0	4687	1 PLE1_RAT	P30427 rattus norv

RESULT 1				
TPR_HUMAN				
ID	TPR_HUMAN	STANDARD;	PRT;	2349 AA.
AC	P12270;			
DT	01-OCT-1989	(Rel. 12, Created)		
DT	01-OCT-1996	(Rel. 34, Last sequence update)		
DT	15-MAR-2004	(Rel. 43, Last annotation update)		
DE	Nucleoprotein TPR.			
GN	TPR.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93064711; PubMed=1437155;			
RA	Mitchell P.J., Cooper C.S.;			
RT	"The human tpr gene encodes a protein of 2094 amino acids that has			
RT	extensive coiled-coil regions and an acidic C-terminal domain.";			
RL	Oncogene 7:2329-2333(1992).			
RN	[2]			
RP	REVISIONS, AND CHARACTERIZATION.			
RX	MEDLINE=95096166; PubMed=7798308;			
RA	Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,			
RA	Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;			
RT	"Tpr, a large coiled coil protein whose amino terminus is involved in			
RT	activation of oncogenic kinases, is localized to the cytoplasmic			
RT	surface of the nuclear pore complex.";			
RL	J. Cell Biol. 127:1515-1526(1994).			
RN	[3]			
RP	SEQUENCE OF 1-142 FROM N.A.			
RX	MEDLINE=88262257; PubMed=3387099;			
RA	King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;			
RT	"tpr homologues activate met and raf.";			
RL	Oncogene 2:617-619(1988).			
CC	-!- FUNCTION: Component of the cytoplasmic fibrils of the nuclear pore			
CC	complex implicated in nuclear protein import. Its amino terminus			
CC	is involved in activation of oncogenic kinases.			
CC	-!- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE			
CC	COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH			
CC	TRP-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER			
CC	COMPONENTS, INCLUDING P62.			
CC	-!- TISSUE SPECIFICITY: Highest in testis, lung, thymus, spleen and			
CC	brain, lower levels in heart, liver and kidney.			
CC	-!- DISEASE: Involved in tumorigenic rearrangements with the MET, TRK			
CC	or RAF genes.			
CC	-!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;			
CC	WWW="http://www.infobiogen.fr/services/chromcancer/Genes/TPRID282.html".			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			

Q9ukx3 homo sapien
P02565 gallus gall
Q13439 homo sapien
Q9ukx2 homo sapien
Q14789 homo sapien
O67124 aquifex aeo
Q99996 h a-kinase
Q96yr5 sulfolobus
P02567 caenorhabdi
Q15149 homo sapien
P79293 sus scrofa
Q28641 oryctolagus

34 426.5 7.0 1938 1 MYHD_HUMAN
35 426 7.0 1940 1 MYH3_CHICK
36 425.5 7.0 2230 1 GOA4_HUMAN
37 425 7.0 1941 1 MYH2_HUMAN
38 422 6.9 3259 1 GOB1_HUMAN
39 420 6.9 978 1 RA50_AQUAE
40 420 6.9 3911 1 AKA9_HUMAN
41 417 6.9 879 1 RA50_SULTO
42 417 6.9 1938 1 MYSD_CAEEL
43 414 6.8 4684 1 PLE1_HUMAN
44 413 6.8 1935 1 MYH7_PIG
45 413 6.8 1938 1 MYH4_RABIT

ALIGNMENTS

CC EMBL; X66397; CAA47021.1; -;
DR EMBL; Y00672; CAA68681.1; -;
DR Genew; HGNC:12017; TPR.
DR MIM; 189940; -;
DR GO; GO:0005737; C:cytoplasm; TAS.
DR GO; GO:0005643; C:nuclear pore; TAS.
DR GO; GO:0006606; P:protein-nucleus import; TAS.
KW Coiled coil; Proto-oncogene; Chromosomal translocation;
KW Nuclear protein; Transport.
FT DOMAIN 78 360 COILED COIL (POTENTIAL).
FT DOMAIN 422 571 COILED COIL (POTENTIAL).
FT DOMAIN 575 628 COILED COIL (POTENTIAL).
FT DOMAIN 758 805 COILED COIL (POTENTIAL).
FT DOMAIN 834 869 COILED COIL (POTENTIAL).
FT DOMAIN 934 979 COILED COIL (POTENTIAL).
FT DOMAIN 1004 1064 COILED COIL (POTENTIAL).
FT DOMAIN 1138 1166 COILED COIL (POTENTIAL).
FT DOMAIN 1196 1241 COILED COIL (POTENTIAL).
FT DOMAIN 1262 1304 COILED COIL (POTENTIAL).
FT DOMAIN 1354 1434 COILED COIL (POTENTIAL).
FT DOMAIN 1476 1595 COILED COIL (POTENTIAL).
FT DOMAIN 527 530 POLY-SER.
FT DOMAIN 1833 1836 POLY-GLU.
FT DOMAIN 1957 1964 POLY-ASP.
FT DOMAIN 2295 2298 POLY-SER.
SQ SEQUENCE 2349 AA; 265600 MW; AFDD6885CEDCA9EF CRC64;

Query Match 7.8%; Score 476; DB 1; Length 2349;
Best Local Similarity 20.4%; Pred. No. 1.4e-08;
Matches 314; Conservative 252; Mismatches 510; Indels 460; Gaps 67;
QY 34 EDAKKNKANRKEE---DWMASG-TIKRH-----LKPSEGESEKKT 68
DB 508 EARGNHVIRDEEVSSADSSSEVISQHLVSRYNIEELQQNQRLVALRELGETRERE 567
QY 69 KKSVELSK-----EDLIQLLSIMEGELQAREDAVI-----HMLRTEKTKPEVLE 111
DB 568 EQETTSKITELQLKLESALTELEQLRKRQHQMQLVDSIVRQDMYRILLSQTTGVAIP 627
QY 112 AHYGSAE-----PEKVL---RVLHRDAIIAQEKSIGEDVYEKPI 147
DB 628 LHASSLDVSLASTPKRPSTSQTVSTPAPVPVIESTEAEAKAALKQLQEIFEN-YKKEK 686
QY 148 SELDRLEBKQKETR-----RMLEQLLLAEK-----CHRRTVYELENEK 186
DB 687 AENEKIQNEQLEKLEQVTDLRSONTKISTQDFASKRYEMLQDNVEGYRREITSLHERN 746
QY 187 HKHTDYMNKSDDFTNLLEQE-RERLKKLLEQE-KAYQARKEKENAK-----RLNKLDELV 240
DB 747 QKLTATQKQEQIINTMTQDLRGANEKLAVAEVRANLKKKEMLKLSVRLSQRESL- 805
QY 241 KLKSFALMLVDERQMHIEQLGLQS-----QKVQDLTQKLREEEKLVKAVTYK 287
DB 806 -----LAEQRGQNLLLTNLQTIQILRSETETQRLSSQIEKLEHEISHLKKLEN 857
QY 288 SKEDRQKLLK-LEVD-FEHKASRFSQBHEEMNAK-----LANQESHNRQLRLKLVG 336
DB 858 EVEQRHTLTRNLDVQLLDTKRQLDTETNLHLNTKELLKNAQKEIATLKQHLNMEVQVAS 917
QY 337 -LSQRI-----EELBETNKSLOKABEELQELREKIAKGECSNLSMAEVESLKRVL 387
DB 918 QSSQRTGKGQPSNKEDVDDLVSQLRQTEEQVNDLKERLKTSTSNVEYQAMVTSLESLN 977
QY 388 EMEGKDBEITKT-EAQCRE-----LKKKLQEEHHHSKEL-----RLEVEKLQKRMSLE 435
DB 978 KEKQVTEEVVRKNIIEVRLKESAEFQTQLEKKLMEVEKEKQELQDDKRRRAIESMEQQLSELK 1037
QY 436 K-----LEEAFSR-----SKSEC-TQLHLNLEKEKNLTKDLL-----NELE 470
DB 1038 KTLSSVQNEVQEQALQRASTALSNEQQARRDCQEQAKIAVEAQNKYVERELMLHAADVEALQ 1097
QY 471 VVKSRVKE-----LECSESRLKEAELSLLKDDLTGK-----LKSFT 504

DB 1098 AAKEQVSKMASVRQHLBETTQKAESQLLECKAS-WEERERMLKDEVSKVCRCEDLEKQN 1156
QY 505 VMLVDERKNMMEKI KQBERK-VDG-LNKNFKVE---QGKMDV-----TEKLIIEESK--- 551
DB 1157 RLLHDQIEKLSDKVVASVKEGVQGPLNVSLSEEGKSQSOILEILFRIFIREKEIAETRFEV 1216
QY 552 -----KLLKLKSEMEEKEYSLTKER-----DELMGK----- 577
DB 1217 AQVESLRYRQRVRELLERLELEDSLNAERERKQVQVTAQTMAQHEELMKKTETMNVVMEIN 1276
QY 578 --LRSEER-----SCILSCSVDLL---KKRLDGIEEVER 607
DB 1277 KMLREKERLEQDLQOMQAKVRKLELDILPLQEANABELSEKSGMLQAEKKLLE--EDVKR 1334
QY 608 EINRGSKSGSEFTCPEDNKKIRELTLEIERLKKRLQQL----- 645
DB 1335 WKARNQHLV-SQOKDPDTEEYRKLLSEKEVHTKRIQQLTEEIGRLKAEIARSNASLTNNQ 1393
QY 646 --EVVEGDLMKTEDEYD---QLEOKFRTEQDKANFLSQ-----OLEEIKHQMA 688
DB 1394 NLIQSLKEDLNKVRTEKETIQKDLDAKIIDIQEKVKTITQVKKIGIRRYKTQYEELKAQOD 1453
QY 689 K-----HKAIEKGEAVSQEAELRHRFRLEEARSDLOAEVQ---ALKEKIHLEL 733
DB 1454 KVMETSAQSSGDHQ--EQHVSVOEQMQLKETLNAETKSKSLESQVENLQKTLSEKETE 1511
QY 734 MNKEDQLSOLQVDYSLVLOQRFMEETKKNMKGREVLNLTKELELSKRYSRALRPSGNRR 793
DB 1512 RNLQEQTVQLQSELRLRQDLQDRTTQEEQLRQOI---TEKEEKTRK----- 1555
QY 794 MVDVPEVASTGVQTEAVCGDAEEETPAVFIRKSFQEBENHIMSNLRQVGLKKPME-RSSVL 852
DB 1556 --AIVAASKIAHLAGVKDQLTKENEELKQRNGALDQKDELQDVRITALKSQYEGRISRL 1613
QY 853 DRYPPAANELTMRKSWIPMRKRENGPSTPOEKGPRPNQAGHPGELVLAPKQGQPLHIR 912
DB 1614 ER-----ELREHQRHLEQRDE-POEPSNK-----VPEQQRQITLK 1648
QY 913 VTP-----DHENSTATLEITSPTSEFFSFTVIPTLGNQK-PRITIIIPSPNVMSQ 962
DB 1649 TTPASGERGIASTSDPPTANIKPTPVVSTP--SKVTAAAMAGNKSSTPRASIRP---MVT 1702
QY 963 KPXSADPTLGPERRAMPVTITTSREKSPESGERSAFADRPASPIQIMTVSTSAAPTEIAV 1022
DB 1703 PATVTNPTTTPTATVMTPTTQVESQEQAMQSEG-----PVEHVPVFGSTSGSVRST 1751
QY 1023 SPESQEVPMGRITILKVTPEKQTVAPVRKYNSNANIITTEDNKIHLHLSQFKRSPGPAA 1082
DB 1752 SPNVQP-SISQPILTVQQQTQ-----ATAFVQPTQQSHPOIEPANQ-----ELS 1794
QY 1083 EGVSPVITVRPVNVTAKEVSTGTGLRSPRNHLSSRPGASKVTSTI----- 1128
DB 1795 SNIVEWVQSSPVERPSTSTAVFGTVSATPSSSLPKRTRREEEDSTIEASDQVSDDTVEMP 1854
QY 1129 -----TITPV-TTSSSTRGTQSVSGQDSSQRPPTPTTRIPMSKGMKAGKPVVAASGAGNLT 1181
DB 1855 LPKKLKSVPVGTGTEEEVMAEESTDGEVETQVYNQDSQDSIGEGVT-----QGDT 1904
QY 1182 KFQPRAE-TQSMKIE---LKKSAASSTASLGG-CKG 1212
DB 1905 PMEDSEETSQSLQIDLGLPIQSDQQTITSSQDQGGKG 1940

RESULT 2
MYH9 HUMAN
ID MYH9 HUMAN STANDARD; PRT; 1960 AA.
AC P35579; O60805;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,
type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).

GN MYH9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Copley V.E., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Evans P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA Gilbert J.G.R., Fey J.M., Fleming K., French L., Garner A.A.,
RA Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA Mcclell J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spragon L., Steward C.A., Suleton J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
RA Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA Khan A.S., Lane L., Tilahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RN [2]
RP SEQUENCE OF 1-1337 FROM N.A.
RX MEDLINE=92003925; PubMed=1912569;
RA Toothaker L.E., Gonzalez D.A., Tung N., Lemons R.S., le Beau M.M.,
RA Arnaout M.A., Clayton L.K., Tenen D.G.;
RT "Cellular myosin heavy chain in human leukocytes: isolation of 5'
RT cDNA clones, characterization of the protein, chromosomal
RT localization, and upregulation during myeloid differentiation.";
RL Blood 78:1826-1833(1991).
RN [3]
RP SEQUENCE OF 1-715 FROM N.A.
RX MEDLINE=91316803; PubMed=1860190;
RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,
RA Gdula D., Adelstein R.S., Weir L.;
RT "Human nonmuscle myosin heavy chains are encoded by two genes located
RT on different chromosomes.";
RL Circ. Res. 69:530-539(1991).
RN [4]
RP SEQUENCE OF 714-1960 FROM N.A.
RX MEDLINE=90138958; PubMed=1967836;
RA Saez C.G., Myers J.C., Shows T.B., Leinwand L.A.;
RT "Human nonmuscle myosin heavy chain mRNA: generation of diversity

through alternative polyadenylation.";
Proc. Natl. Acad. Sci. U.S.A. 87:1164-1168(1990).
[5]
RP VARIANT DFNA17 HIS-705.
RX MEDLINE=20489856; PubMed=11023810;
RA Lalwani A.K., Goldstein J.A., Kelley M.J., Luxford W., Castelein C.M.,
RA Mhatre A.N.;
RT "Human nonsyndromic hereditary deafness DFNA17 is due to a mutation in
RT nonmuscle myosin MYH9.";
RL Am. J. Hum. Genet. 67:1121-1128(2000).
RN [6]
RP VARIANTS MHA/FTNS/SBS LYS-93; CYS-702; CYS-1165; HIS-1424 AND
RP LYS-1841.
RX MEDLINE=20428192; PubMed=10973259;
RA Seri M., Cusano M., Gangarossa S., Caridi G., Bordo D., Lo Nigro C.,
RA Ghiggeri G.M., Ravazzolo R., Savino M., Del Vecchio M., d'Apolito M.,
RA Iolascon A., Zelante L.L., Savoia A., Balduini C.L., Noris P.,
RA Magrini U., Belletti S., Heath K.E., Babcock M., Glucksman M.J.,
RA Aliprandis E., Bizzaro N., Desnick R.J., Martignetti J.A.;
RT "Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and
RT Sebastian syndromes.";
RL Nat. Genet. 26:103-105(2000).
RN [7]
RP VARIANTS MHA ILE-1155 AND LYS-1841.
RX MEDLINE=20428193; PubMed=10973260;
RA Kelley M.J., Jawien W., Ortel T.L., Korczak J.F.;
RT "Mutation of MYH9, encoding non-muscle myosin heavy chain A, in
RT May-Hegglin anomaly.";
RL Nat. Genet. 26:106-108(2000).
CC -!- FUNCTION: Cellular myosin appears to play a role in cytokinesis,
CC cell shape, and specialized functions such as secretion and
CC capping.
CC -!- SUBUNIT: Myosin is a hexameric protein that consists of 2 heavy
CC chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2
CC regulatory light chain subunits (MLC-2).
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC -!- DISEASE: Defects in MYH9 are the cause of May-Hegglin anomaly
CC (MHA) [MIM:155100]. MHA is an autosomal dominant
CC macrothrombocytopenia characterized by thrombocytopenia, giant
CC platelets and leukocyte inclusions.
CC -!- DISEASE: Defects in MYH9 are the cause of Fechtner syndrome (FTNS)
CC [MIM:153640]. FTNS is an autosomal dominant macrothrombocytopenia
CC characterized by thrombocytopenia, giant platelets and leukocyte
CC inclusions. With additional alport-like clinical features of
CC sensorineural deafness, cataracts and nephritis.
CC -!- DISEASE: Defects in MYH9 are the cause of Sebastian syndrome (SBS)
CC [MIM:605249]. SBS is an autosomal dominant macrothrombocytopenia
CC characterized by thrombocytopenia, giant platelets and leukocyte
CC inclusions.
CC -!- DISEASE: Defects in MYH9 are the cause of autosomal dominant
CC nonsyndromic sensorineural deafness 17 (DFNA17) [MIM:603622].
CC DFNA17 is characterized by progressive hearing impairment and
CC cochleosaccular degeneration.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z82215; -; NOT ANNOTATED_CDS.
DR EMBL; M81105; AAA59888.1; -.
DR EMBL; M69180; AAA61765.1; -.
DR EMBL; M31013; AAA36349.1; -.
DR PIR; A61231; A61231.
DR HSSP; P10587; 1BR2.
DR Genew; HGNC:7579; MYH9.

DR	MIM; 160775; --	
DR	MIM; 155100; --	
DR	MIM; 153640; --	
DR	MIM; 605249; --	
DR	MIM; 603622; --	
DR	GO; GO:0005860; C:non-muscle myosin; TAS.	
DR	InterPro; IPR000048; IQ region.	
DR	InterPro; IPR001609; myosin_head.	
DR	InterPro; IPR004009; Myosin_N.	
DR	InterPro; IPR002928; Myosin_tail.	
DR	InterPro; IPR002017; Spectrin.	
DR	Pfam; PF00612; IQ; 1.	
DR	Pfam; PF00063; myosin_head; 1.	
DR	Pfam; PF02736; Myosin_N; 1.	
DR	Pfam; PF01576; Myosin_tail; 1.	
DR	PRINTS; PR00193; MYOSINHEAVY.	
DR	ProDom; PD000355; myosin_head; 1.	
DR	SMART; SM00015; IQ; 1.	
DR	SMART; SM00242; MYSc; 1.	
DR	PROSITE; PS50096; IQ; 1.	
KW	Myosin; ATP-binding; Calmodulin-binding; Actin-binding; Coiled coil;	
KW	Multigene family; Disease mutation; Deafness.	
FT	DOMAIN 1 778 MYOSIN HEAD-LIKE.	
FT	DOMAIN 779 808 IQ.	
FT	DOMAIN 837 1926 COILED COIL (POTENTIAL).	
FT	NP_BIND 174 181 ATP (POTENTIAL).	
FT	DOMAIN 654 676 ACTIN-BINDING.	
FT	VARIANT 93 93 N -> K (IN MHA).	
FT	702 702 /FTid=VAR 010791.	
FT	705 705 R -> H (IN DFNA17).	
FT	1155 1155 /FTid=VAR 010793.	
FT	1165 1165 /FTid=VAR 010794.	
FT	1424 1424 R -> C (IN SBS).	
FT	1841 1841 D -> H (IN FTNS).	
FT	53 55 /FTid=VAR 010796.	
FT	660 660 E -> K (IN MHA).	
FT	869 869 /FTid=VAR 010797.	
FT	931 931 EAI -> RGH (IN REF. 3).	
FT	1240 1241 T -> S (IN REF. 3).	
FT	1350 1350 T -> M (IN REF. 4).	
FT	CONFLICT 53 55 C -> Y (IN REF. 4).	
FT	CONFLICT 660 660 KG -> GR (IN REF. 4).	
FT	CONFLICT 869 869 E -> EE (IN REF. 2).	
FT	CONFLICT 931 931	
FT	CONFLICT 1240 1241	
FT	CONFLICT 1350 1350	
Query Match		
Best Local Similarity 7.7%; Score 469; DB 1; Length 1960;		
Matches 228; Conservative 179; Mismatches 343; Indels 246; Gaps 40;		
QY	67 KTKKSVELSKEDLIQLLSIMEGELQAREDVIHMLRTEKTPKEVLEAHYGSAPKVLRLV 126	
Db	833 KVKPLLVSRQ-----EEMWAKEEEL-----VKVREKQLAAENRLTEMETLQSQL 878	
QY	127 HRDAILAQEKSIGEDVYEKPISEL-DRLEEKQKETYRRMLEQLLLAEKCH-RRTVYELEN 184	
Db	879 MAEKLQLEQLQAE TELCAEAEELRARLTAKQE-----LEEICHDLEARVVEEEE 928	
QY	185 EKHKHTDYMNKSDFTNLLLEQERERLKKLLEQEKAVQARKEKENA-----KRLNKLDELVK 241	
Db	929 ERCQHLQAEKK-----MQQNIQEELEEEESAPQKLQLEKVTTEAKLKKLEEEQII 982	
QY	242 LKSFALMLVDERQMHIQGLQSQKQVLDLTQKLREEBEKLVAVTYKSKEDRQKLLKLEVD 301	
Db	983 LEDQNCKLAKEKK-----LLEDRIAFTTNLTEEBEKSLSA-KLKNKHEAMI---TD 1031	
QY	302 FEHKASRFSQHEEEMNAKLANQESHNPQLRLKLVGLSORIEELE-----ETNKSLOKAEEE 357	
Db	1032 LEERLREEKQROEL-----EKTRKLEGDSTDLSQIAELQAIQIAELKMQLAKKEEE 1084	
QY	358 LQELREKIAKGECCGSSLMAEVESLRKRVLEME-----GKDBEIT 397	

Db	1085 LQAALARVEEEAAQKNMALKKIRELESQISELQEDLESESRASNKAQKRDLGEELEAL 1144	
QY	398 KTE-----AQCRELK-----KKLQEEHHHSEKRLVEKIQKRMSELEKLEE 439	
Db	1145 KTELEDTLDSTAAQQELRSKREQEVNLIKKTLEEEAKTHEAQIQ-EMROKHSQAVEELAE 1203	
QY	440 AFSRSKSECTQLHLNLEKEK----NLT KDLLNELEV-----VKS RVKELEC 481	
Db	1204 QLEQTK---RVKANLEKAKQTLNENERGELANEVKVLLQGGKGDSEHKRKVKVEAQLQELQV 1259	
QY	482 SESRLEKAELSLKDDLTKLKSFTVMLVDERKNMMEKIQBEERKVDGGLNKNFKVQGGKVM 541	
Db	1260 KFNEGERVRTTELADKVTKLQV-----ELDNVTGLLSQSDSKSSKLT KDFSALSQLQD 1312	
QY	542 VTEKLIEESKLLKLLKSEMEEKEYSLTKERDELMGKLRSEERSCV-----D 593	
Db	1313 TOELLQENROKLSLSFKLKQVE----DEKNSFREQUELEEEEAHNLEKQIATLHAQVAD 1368	
QY	594 LLKKRLDGI-----BEVEREINR---GRSCKGSEFTCPEDNKKIRELTLEIERLKKRLQ 644	
Db	1369 MKKIMEDSVGCLETAEEVKRKLDLEGLSQRHEEKVAAVD-----KLEKTKTRLQ 1420	
QY	645 -LE--VVEGD-----LMKTEDEYDQL-----BQKFRTEQDKANFLSQLEEEIKHQ 686	
Db	1421 ELDDLVDLDHQRQACNLEKKQKQKFDQLLAEKTTISAKYAEERDRAEAEAREKETKALS 1480	
QY	687 MAKHKAIEKGEAVSQEAELRH---RFRLE-----EAKSRDLQAEVQA 725	
Db	1481 LA--RALE--EAMEQKAELERLNKQFRTMEDLMSSKDDVGVKS VHELEKSKRALEQQVVEE 1536	
QY	726 LKEKIHLMNK-----EDQLSQLQVDYSV-----LQQRFMEEETKNKNMGREVLNLT 773	
Db	1537 MKTQLEEELEQATEDAKLRLEVLNQLAMKAQFERDLQGRDEQSEKKKQLVQRVREMEA 1596	
QY	774 ELELSKRYSRALRPSGNGRRMVDVPVASTGVQTEAVCGDAAEEETPAVFIRKSFQEE--- 830	
Db	1597 ELE-DEKQORSMAVARKKLEMDLK-----DLEAHIDSANKNRDEAIKQLRKLOAQMKD 1649	
QY	831 -----NHIMSNLRQVGLK-KPMERSSVLD RYPPAANELTMRKSWIPWMRKREN 877	
Db	1650 CMRELD DTRASREEILAQAKENEEKLKSMEAEMIQLQEEELAAAEARAKQA----- 1699	
QY	878 GPSTPQEKGRPRNQAGHPGELVLPAPKQGQPLHIRV 913	
Db	1700 ----QQRDELADEIANSSGKGALALEEKKRRLEARI 1731	
RESULT 3		
CENE_HUMAN		
ID	CENE_HUMAN	STANDARD; PRT; 2663 AA.
AC	Q02224;	
DT	01-JUL-1993 (Rel. 26, Created)	
DT	01-JUL-1993 (Rel. 26, Last sequence update)	
DT	10-OCT-2003 (Rel. 42, Last annotation update)	
DE	Centromeric protein E (CENP-E protein).	
GN	CENPE.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=93024922; PubMed=1406971;	
RA	Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;	
RT	"CENP-E is a putative kinetochore motor that accumulates just before	
RT	mitosis."	
RL	Nature 359:536-539 (1992).	
RN	[2]	
RP	CHARACTERIZATION.	
RX	MEDLINE=95196755; PubMed=7889940;	
RA	Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;	
RT	"Mitotic Hela cells contain a CENP-E-associated minus end-directed	

RT microtubule motor.";
RN EMBO J. 14:918-926(1995).
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=97634420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
RN [4]
RP FARNESYLATION.
RX MEDLINE=20459117; PubMed=10852915;
RA Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,
RA Bishop W.R., Kirschmeier P.;
RT "Farnesyl transferase inhibitors block the farnesylation of CENP-E
RT and CENP-F and alter the association of CENP-E with the
RT microtubules.";
RL J. Biol. Chem. 275:30451-30457(2000).
CC -I- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -I- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -I- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -I- SIMILARITY: Belongs to the kinesin-like protein family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z15005; CAA78727.1; -.
DR PIR; S28261; S28261.
DR HSSP; P17119; 3KAR.
DR Genew; HGNC:1856; CENPE.
DR GK; Q02224; -.
DR MIM; 117143; -.
DR GO; GO:0005699; C:kinetochore; TAS.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0008350; F:kinetochore motor activity; TAS.
DR GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
DR GO; GO:0007079; P:mitotic chromosome movement; TAS.
DR GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere; Lipoprotein; Prenylation.
FT DOMAIN 1 335 KINESIN-MOTOR.
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP BIND 86 93 ATP (BY SIMILARITY).
FT LIPID 2660 2660 S-farnesyl cysteine.
SQ SEQUENCE 2663 AA; 312087 MW; CEFC13880C8C8C8 CRC64;

Query Match 7.7%; Score 469; DB 1; Length 2663;
Best Local Similarity 20.3%; Pred. No. 2.6e-08;
Matches 279; Conservative 267; Mismatches 455; Indels 376; Gaps 62;

QY 33 SEDAKONKANRKEED-----VNASGTIKRHLKPSGESEKTK-----KSVE 73
DB 1382 SQSKQEQSLNMBKDNETTIVSEMEQFKPKDSALLRIEIMGLSKRLQESHDEMKVA 1441

QY 74 LSKEDLIQLLSIMEGEL-QAREDVHMLRTE-KTKPEVLEAHYGSAPAEKVLRLVLRDAI 131
DB 1382 SQSKQEQSLNMBKDNETTIVSEMEQFKPKDSALLRIEIMGLSKRLQESHDEMKVA 1441

Db 1442 KKKDDLRQLQVLOSQSDQLKENIKIIVAKHLETEELKVAHCCLKBQEETINELR---V 1498
QY 132 LAQEKSIGEDVVEKPISEL-DRLEEKQKQETVRRMLEQQLLAE-----KCHR 176
Db 1499 NLSEKETEISTIQKLEAINDKLNKIQEIYEKE-EQLNIKQISEVQENVNELKQFKEHR 1557
QY 177 R---TVYELENEKHKHTDYMNKSDDFTNLLEQERERLKKLLEQEKAYQARKE--KENAK 230
Db 1558 KAKDSALQSIEMKMLTNRQLQESQEEIQIMIKESMKRV---QEALQIERDQLKENTK 1614
QY 231 RL-NKLRDELVK---LKSFALMLVDERQMHIEQL--GLQSQKVQ-----DLTQKL 274
Db 1615 EIVAKMKESQEKYQFLKMTAVNETQEKMCETIEHLKEQFETQKLNLENIENTENIRLTQIL 1674
QY 275 REEEKLVKAVTYKSKEDR---QKLLKLEVD-----FEHKASRFSQEHHEEMNAKLANQESH 326
Db 1675 HENLEEMRSVT-KERDDLRSVEETLKVERDQLENLRETITRDLEKQEEELKIVHMLKEH 1733
QY 327 NRQLRLKLVGL-SQRIEELEETNKSLOKABEELQ---ELREKIAKGECGSSSLMAEVES 381
Db 1734 QETID-KLRGIVSEKTNISNMQKOLEHSNDALKAQDLKIQEELRIAHMHLKBQETIDK 1792
QY 382 LRKRVLEMEGK-----DEEITKTEAQACRELKKKLQEEHHSHKELRLEVEKLOKRMSELEKL 437
Db 1793 LRGIKSEKTLKSNMQKDLNSNAKLOEKIQELKANEHQLITLKQDVNETQKVSEMEQL 1852
QY 438 ----EEAFSRKSECTQLHLNLEKEKNLTLDLNELEVVKSRVKE---LECSERLEKA 489
Db 1853 KKQIKDQSLTSLKLEIENL-----NLAQELHENLEEMKSVMKERDNLRRVEETLKLE 1904
QY 490 ELSLKDDLTKLKSFTVMLVDERKNMMEKIKQEEKVVDGLNKNFKVEQGKVMVDVTEKLIIE 549
Db 1905 RDQLKESLQETKARDLEIQOELKTARMLSKHEKTVDKLRE-----KISEKTIQISDI 1957
QY 550 SKLLKLVKSEMEKEYSLTKERDELMGKLRSEHEERSCELSVDLLKKRLDGLIEEVEREI 609
Db 1958 QKDLDKSKDELQKKIQELQKKELQL---LRVKED-----VNMSHKKINEMEQKKQF 2006
QY 610 NRGSRCKGSEFTCPEDNKKIRELTLEIERLKKRLLQOLEVVEGDLTKTEYDQLEQKFRTE 669
Db 2007 EPNYLCK-----CEMDN--FOLT---KKLHESLEEIRIV---AKERDELRRIKESLKME 2052
QY 670 QDKANFLSQQLLEEI---KHQMAKHKAI-----EK---GEAVSQEAELRH 707
Db 2053 RDQ--FIATLREMIARDRQNHQVKPEKRLSDGQQHLMESLREKCSRIKELLKRYSEMDD 2110
QY 708 RFRLEEAKSRDLQAEVQ-----ALKEKIHLMNKKEDQLSQLQVDYSVLQ 751
Db 2111 HYECLNRLSLDLEKEIEFHRIMKKLVLSYVTKIKEEQHECINK-----FE 2157
QY 752 QRFMEEEETKNKMGREVLNLTKELELSKRYSRALRPSGNGRRMVDVPVASTGVQTEAVCG 811
Db 2158 MDFIDEVEKQKELLIKIQHLQDQCDVPSRELRLDKLNQN---MDLHI-----EELK 2206
QY 812 DAAEEETPAVFIRKSFQENHIMSRLQV-----GLKKPMERSSVL 852
Db 2207 DFSESEFPS--IKTEFQQ---VLSNRKEMTQFLEEWLNTFRDIEKLKNGIQKENDRICQV 2261
QY 853 -----DRYPPAANE-----LTMKRSWIPWMKRENGPSTPQKGPKNQAGHPGEL 899
Db 2262 NFFNRRITAIMNESTEFERSATISKEW-----EQDLKSLKKEKNEK-----2303
QY 900 VLAPKQGOPLHIRVTPDHENSTATLEITSPTSEFFSSTTVIPTLGNOKPRITIPSPNV 959
Db 2304 --LFGNYQTL-----KTSLASGAGVNPPTQD-----NKNPHVT-SRATQL 2340
QY 960 MSQPKPSADPTL--GPERAMS-----PVTITISREKSPGCRSAFADRPASP 1005
Db 2341 TTEKIRELENSLHEAKESAMHKSIIKMQKELEVTNDIIAKLOAKVHESNKCLEKTKET 2400
QY 1006 IQIMTVTSAAPTEIAVSPESQEV-PMGRTILKVTPEKQTVPAVPRKYNSNANIITTEDN 1064
Db 2401 IQVLQDKVA-----LGAKPYKEEIEDLKMVLKIDLEKMNAKEFEK-EISATKATVEYQ 2454

Qy 1065 KIHHL-----GSQFKRSPGAAGVSPVITVRPV-----NVTA 1098
Db 2455 KEVIRLLRENLRRSQAQDTSEHTDPQPSNKP LTCGGSGIVQNTKALILKSEHRL 2514
Qy 1099 EKEVS-----TGTVLRSPRNHLSSR-----PGASKVT 1125
Db 2515 EKEISKLKQNEQLIKQKNELLSNNQHLNSNEVKTWKRTLKREAHKQVTCENSPKSPKVT 2574
Qy 1126 STIT---ITPVTTSSTRGTQSVSGQDSSQRPTPTTRIP-----MSKGMKAGKPV 1171
Db 2575 GTASKKKQITPSQC-----KERNLQDPVPKESPKSCFFDSRSKSLPSPHPV 2620

RESULT 4
USO1_YEAST
ID USO1_YEAST STANDARD; PRT; 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Intracellular protein transport protein USO1.
GN USO1 OR INT1 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, uso1, is required for intracellular
protein transport in Saccharomyces cerevisiae.";
RL J. Cell Biol. 113:245-260(1991).
RN [2]
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]

RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Required for protein transport from the ER to the Golgi
complex.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
ER AND THE GOLGI COMPLEX.
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, composed
of an heptapeptide repeat pattern characteristic of alpha-helical
coiled coils. May form filamentous structures in the cell.
CC -!- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----
CC EMBL; X54378; CAA38253.1; -.
DR EMBL; L03188; AAB00143.1; -.
DR EMBL; U53668; AAB66659.1; -.
DR GermOnline; 140300; -.
DR SGD; S0002216; USO1.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR006955; USO1_p115_C.
DR InterPro; IPR006953; USO1_p115_head.
DR Pfam; PF04871; USO1_p115_C; 1.

DR Pfam; PF04869; USO1_p115_head; 1.
KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724 GLOBULAR HEAD.
FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 847 847 G -> E (IN REF. 2).
FT CONFLICT 924 924 E -> K (IN REF. 2).
FT CONFLICT 1253 1253 V -> I (IN REF. 2).
FT CONFLICT 1319 1319 I -> V (IN REF. 2).
FT CONFLICT 1461 1461 N -> S (IN REF. 2).
FT CONFLICT 1581 1581 G -> S (IN REF. 2).
FT CONFLICT 1600 1600 I -> V (IN REF. 2).
FT CONFLICT 1661 1661 R -> S (IN REF. 2).
FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

Query Match 7.6%; Score 464.5; DB 1; Length 1790;
Best Local Similarity 22.4%; Pred. No. 2.5e-08;
Matches 203; Conservative 184; Mismatches 322; Indels 199; Gaps 32;
Qy 34 EDAKKNKANRKE-----EDVMASGTTIKRHLKPSGESEKTKKSVLSKEDLIQLLSIMEG 88
Db 889 EKDKSNVNHQKETKSLKEDIAAKITEIKAINENLEEMKIQCNNLSKEKEHISKELVEYKS 948
Qy 89 ELQAREDVIHMLRTEKTPPEVLEAHYGSAPPEKVLRLHRDAILAQEKSIGEDVYEKPI 148
Db 949 RFQSHDNLVAKL-TEKLKSLANNYKDMAENESLIKAVEES---KNESSIQLSNLQNKID 1004
Qy 149 ELDRLEEK---QKETVRRMLEQLLLAEKCHRRTVVELENEKHGHTDYNMKSDDFTNLL 205
Db 1005 SMSQKENFQIERGSIETKNIEQL-----KKTISDLQETKE---EIISKSDSSKDEYES 1054
Qy 206 ERERLKKLLEQEKAVQARKEKEN-AKRLNKLRLDELVKLKSFALMLVDERQMHIQGLQ 264
Db 1055 QISLLKEKL--ETATTANDENVNKISLTKTREETEAELAAYKNLKNELETKLETSEK 1112
Qy 265 QKVODLTQKLREEEKKLAVTYKSKEDRQKLLKLEVDPEHKASRFSQHEEMNAKLANOE 324
Db 1113 KEVKEEHLKEEKIQLEK---EATETKQQLNSLRANLE---SLEKEHEDLAAQLKKYE 1165
Qy 325 ----SHNRQLRLKLVLSQRIEELEETNKSLOKAEELQ----- 359
Db 1166 EQIANKERQYNEEISQLNDEITSTQQENESIKKKNDELEGEVKAMKSTSEEQNLKKSEI 1225
Qy 360 ----ELREKIAKGECGSSSLMAEVESLRKRVLEMBGKDEEITKTEAQCRLEKKLQ-EE 413
Db 1226 DALNLQIKELKKKNETNEASLLESIKSVESETVTKELQDECNFKKEVSELEDKLRASE 1285
Qy 414 EHSK-----ELRLEVEKL-----QKRMSELEKL----- 437
Db 1286 DKNSKYLELOKESEKIKEELDAKTTTELKIQLEKITNLKAKEKSESELSRLKKTSSSEERK 1345
Qy 438 --EEAFSRSKSECTQLHLNLEKEKNLTKD-----LLNELEVVK 473
Db 1346 NAEQLEKLKNEIQIKNQAFEKERKLLNEGSSITTOYSEKINTLEDELIRLQENEL-- 1403
Qy 474 SRVKEECSESRLKAELSLKODLTKLKSFTV-MLVDERKNMMEKIQEERKVDGLNKNF 532
Db 1404 -KAKEIDNTRSELEKVSLS-NDELLBEKQNTIKSLQDEILSYKDKITRNDKLLSIERDN 1461
Qy 533 K-----VEQGVMDVTEKLI EESKLLKLKSEMEKEYSLTKERDELMGLRSE-EERS 585
Db 1462 KRDLLESLKEQLRAAQESKAKVEEGLKLEEESSKEKAELEKSK---EMMKLESTIESNE 1518
Qy 586 CELSCSVDLLKKRLDGI EEVEREINRGRCKSGSEFTCPED-----NKIRELTJL 633
Db 1519 TELKSSMETIRKSDKLEQ-----SKSAEEDIKNLQHEKSDLSIRINESEK 1565
Qy 634 EIERLKKRLQOLEVVEGDLTKT-EDEYDQLEQKFRTEQDKANFLSQOLEEIKHQMAKHKA 692
Db 1566 DIEELKSKL-RIEAKSGSELETYVKQELNNAQEKIRINAEENTVLKSKLEDIERELKDKQA 1624

QY 693 IEKGEAVSQEALRRFRLEAEAKSRDLQAEVQALKEKIHLMNKD--QLSQLOVDYSV 749
Db 1625 EIKSNQEEKE-----LLTSLKELEQELDSTQQAQK--SEEERRAEVRKFQVEKSQ 1674
QY 750 LQQRFMEEETK-----NKNMG-----REVNLTKLELSKRYSRALRPS 788
Db 1675 LDEKAMLLETKYNDLVNKEQAKRDEDTVKKTTDSQRQIEKLAKELDNLKAENSKLKEA 1734
QY 789 GNGRRMVD 796
Db 1735 NEDRSEID 1742

RESULT 5
MYH9_RAT
ID MYH9_RAT STANDARD; PRT; 1961 AA.
AC Q62812;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain, type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
GN MYH9.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RA Choi O.H., Park C., Itoh K., Adelstein R.S., Beaven M.A.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Cellular myosin appears to play a role in cytokinesis, cell shape, and specialized functions such as secretion and capping.
CC -!- SUBUNIT: Myosin is a hexameric protein that consists of 2 heavy chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2 regulatory light chain subunits (MLC-2).
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing cycles of a 28-residue repeat pattern composed of 4 heptapeptides, characteristic for alpha-helical coiled coils.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U31463; AAA74950.1; --
CC HSSP; P10587; 1BR2.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Multigene family.
FT DOMAIN 1 778 MYOSIN HEAD-LIKE.
FT DOMAIN 779 808 IQ.
FT DOMAIN 841 1927 COILED COIL (POTENTIAL).
FT NP_BIND 174 181 ATP (POTENTIAL).

FT DOMAIN 654 676 ACTIN-BINDING.
SQ SEQUENCE 1961 AA; 226336 MW; 9B9876D9681FB19E CRC64;
Query Match 7.5%; Score 458.5; DB 1; Length 1961;
Best Local Similarity 23.5%; Pred. No. 4.2e-08;
Matches 195; Conservative 163; Mismatches 283; Indels 189; Gaps 33;
QY 125 VLHRDAILAQEKSIGEDVYEKPISELDRLEEKQKETYRRMLEQLLLAEKCHRRTVYELEN 184
Db 841 IRHEDELLAKEAELTK-VREKHLAAENRLTETMTQSQQLMAEKLQLOEQLOAKTELCAEA 899
QY 185 EKHKHTDYMNKSD-----DFTNLLEQERERUKKLEQEKAYQAR-----KKEENAK 230
Db 900 EELRARTAKQOELEEEICHDLARVEEEERCQYLQAEKKQMQNIQOELEEEESAR 959
QY 231 ---RLNKLKRD--LVKLKSPALMLVDERQMHIEQLGLSQKQVQDLTQKLFREEEKLKAVT 285
Db 960 QKLQLEKVTTEAKLKKEEDQIIMEDQNCKLAKKKLEDRVAEFTTDLMEEEKSKSLA 1019
QY 286 YKSKEDRQKLLKLEVDFEHKASRFSQEHEEMNAKLANQESHNRQLRLKLVGLSQRIEELE 345
Db 1020 -KLKXKHEAMI---TDLEERLRREEKQREL-----EKTRRKLEGGDSTDLSQIAELQ 1068
QY 346 ---ETNKSLOKAEHELQELREKIAKGECSNSSLMAEVESLRKRVLME----- 390
Db 1069 AQIAELKMLAKKEEELQAALARVEEEAAQKNMALKKIRELETQISELQOELESERACRN 1128
QY 391 -----GKDEEITKTE-----AQRELKKKLQEE-----EHSKELRLEVE 425
Db 1129 KAEQKRDLGEELEALKTELEDTLDSTAAQQLRSKREQEVSIKKTLEDEAKTHEAQIQ 1188
QY 426 KL-QKRMSELEKLEEAFFSRKSECTQLHLNLEKEK---NLTKOLLNELEV----- 471
Db 1189 EMRQKHSQAVELEAEQLEQTK---RVKATLEKAKQTLENERGELANEVKALLQKGKDSE 1244
QY 472 ----VKSrvKELECSERLEKAELSLKDDLTCLKSFTVMLVDERKNMMEKIKQEERKVD 526
Db 1245 HKRKKVEAQQLQELQVKFSEGERVTRTELADKVKLQV-----ELDSVTGLNQSDSKSS 1297
QY 527 GLNKNFKVEQGVMDVTEKLIIESKKLLKLMSEMEKEYSLTKERDELGMKLRSEEE--- 583
Db 1298 KLTkDFSALeSOLQDTQELLQENRQKLSLSTKLKQME---DEKNSFRELEEEEEEA 1353
QY 584 RSCELSCS-----VDLLKKRLDGI-----EEVEREINR---GRSCKGSEFTCPEDNKI 628
Db 1354 RNLEKQIATLHAQVTDMMKKMEDGVGCLTAEEAKRRLOKDLGLSQRLEEKVAAAYD--- 1410
QY 629 RELTLEIERLKKRLQQLLEVVE-----GDLMKTEDEYDQL-----EQKFRTE 669
Db 1411 -----KLEKTKTRLQ-Q-ELDDLLVLDLHQRSVSNLEKKQKKQKFDQLLAPEKTI 1464
QY 670 QDKANFLSQOLEEIKHQMAKHKAIEKGEAVSQEAEALRH---RFRLE----- 712
Db 1465 RDRAEAEAREKETKALSIA--RALE--EAMEQKAELERLNKQFRTEMEDLMSSKDDVGKS 1520
QY 713 ----EAKSRDLQAEVQALKEKIHLMNK-----EDQLSQLOVDYSV-----LQQRFME 756
Db 1521 VHELEKSNRALEQQQVEEMKTQLEEELEDEQLQATEDAKLRLEVNQAMKQAFERDLQGRDEQ 1580
QY 757 EETKNKNMGREVLNLTKELELSKRYSRALRPPSGNGRRMVDVPVASTGVQT 806
Db 1581 SEEKKKQLVRQVREMEAELE-DEKQKRSIAMAARKKLEMDLKDLEAHIDT 1629

RESULT 6
MYS_AEQIR
ID MYS_AEQIR STANDARD; PRT; 1938 AA.
AC P24733;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, striated muscle.
OS Aequipecten irradians (Bay scallop).

Db	1570	DKRIQEKEEF	DNTRRHQRALES	MQASLEAEAKG	----	ADAMRIKKKLEQD	INELEVAL	1626	
QY	786	RPSGNGRRM	VDVPVASTGVQ	TEAVCGDAAEE	ETPAVIRKSFQ	-----	EENHIM	834	
Db	1627	DASNRGKAEM	EKTVKRYQQQ	I REMQTSIEE	EQRQREARES	YNMAERRCT	LMSGVEELR	1686	
QY	835	SNLRQVGLK	KPMERSVLD	DRYP	PPAANELTMRK	SWIPMRKKRENGP	-----	STPQEKGRPN	890
Db	1687	AALEQAEAR	KASDNE	LADA-NDRV	NELTSQVSSV	QGQKRKLEGD	INAMQTD	LDENHGEL	1745
QY	891	QGAGHPGEL	VLAPKQGQ	PLHIRVTP	DHEN	919			
Db	1746	KGADERCKK	AMADAAR	LAD	ELRAEQDHSN	1774			

RESULT 7

MYHB_HUMAN STANDARD; PRT; 1972 AA.
AC P35749; O00396; O94944; P78422;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, smooth muscle isoform (SMMHC).
GN MYH11 OR KIAA0866.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99425270; PubMed=10493829;
RA Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
RA Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
RA Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
RT "Genome duplications and other features in 12 Mb of DNA sequence from
RT human chromosome 16p and 16q.";
RL Genomics 60:295-308(1999).
RN [2]
RP SEQUENCE FROM N.A., AND REVISIONS TO 1263-1266.
RC TISSUE=Brain;
RA Nagase T., Kikuno R., Yamakawa H., Ohara O.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-1266 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:355-364(1998).
RN [4]
RP SEQUENCE OF 885-1972 FROM N.A.
RX MEDLINE=93263189; PubMed=7684189;
RA Matsuoka R., Yoshida M.C., Furutani Y., Imamura S., Kanda N.,
RA Yanagisawa M., Masaki T., Takao A.;
RT "Human smooth muscle myosin heavy chain gene mapped to chromosomal
RT region 16q12.";
RL Am. J. Med. Genet. 46:61-67(1993).
RN [5]
RP SEQUENCE OF 1093-1972 FROM N.A.
RC TISSUE=Hippocampus;
RA Okajima K.;
RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Muscle contraction.
CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
CC and 2 regulatory light chain subunits (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- TISSUE SPECIFICITY: Smooth muscle; expressed in the umbilical

Db 840 KVKPLLVTRQ-----EEMQAKEDL-----QKTKERQQAENELKELEQ----- 880
Qy 127 HRDAILAQEKSIGEDVYEKPISELDRLEEKQOKETYRRMLLEQLLAEKCHRRTVVELENEK 186
Db 881 -KHSQLTEEKNL--LQEQLOAQETELYAAEEMRV-----LAACK-----QELEEIL 924
Qy 187 HKHTDYMNSDDFTNLLQEQRER-----LKKLLEQEKAYQARKEKENA---KRLNKLK 236
Db 925 HEMEARLEEDRGQQLQAEARKKMAQQMLDLEEQLLEEAEARQKLQLEKVTAEAKIKKLE 984
Qy 237 DELVKLKSFALMLVDERQMHIHQGLQSGKQVQDLTQKLREBEKLVKAVT----- 285
Db 985 DEI-----LVMDQNNKLSKERKLLLEERISDLTTNLAEBEEKAKNLTCLKNKHESMIS 1037
Qy 286 -----YKSKEDRQKLLKLEVDFEHKASRFESQEHHEEMNAKLANQE---SHNRQLRLKLV 335
Db 1038 ELEVRLKBEKSRQELEKLRKLEGDSDPFHEQIADLQAIQIAELKMQLAKKEEELQAALA 1097
Qy 336 GLSORIELEETNKSLOKABEELQELREKI-----AKGECGNSSSLMAEVESLRKRV-- 386
Db 1098 RLDDIEIAQKNNAKKIRELEGHISDLQEDLDSEARAARNAKAEKQKRDLGEBLEALKTELED 1157
Qy 387 -LEMEGKDEI-TKTEAQCHRELKKLQEE-EHHSKELRLEVEKQLKRMSLEKLEEAFSR 443
Db 1158 TLDSTATQOELRAKREQEVTVLKKALDEETRSHEAQVQEMRQKHAQAVEILTEQEQFKR 1217
Qy 444 SKSECTQLHNLKEKKNLTQDLLNELEV-----VKSrvKELECSSESRLKA 489
Db 1218 AKANLDKNKQTLKE---NADLAGELRVLGOAKQOEVEHKKKKLEAQVQELQSKCSDGERA 1274
Qy 490 ELSLKDDLTKLKSFVTVMLVDERKNMMEKIQEERKVDGLNKNFKVEQGVMDVTEKLEIE 549
Db 1275 RAEKNDKVHKLQ-----NEVESVTGMLNEAEGKAIKLAKDVASLSSQIQDTQELLQEE 1327
Qy 550 SKKLLKLS---EMEEKYSLTKERDELMGKLRSSEERSCELSCSVDLLKKRL----- 599
Db 1328 TRQKLNSTKLRQLEERNLSQDLDEEMAKQNLERHISTLNQLSDSKKKLQDFASTV 1387
Qy 600 DGIEE-----VEREINRGSCKGSEFTCPEDNKIRELTIE 634
Db 1388 EALEEGKKRFQKEIENLTQQYEEKAAAYDKLETKNRLQO-ELDDLVDLDNQ-RQLVSN 1445
Qy 635 IERLKKRLQOLEVEGDL-MKTEDEYDQLEQKFRTEQDKANFLSQOLE--EIKHQMAKH 690
Db 1446 LEKKQKFPDQLLAEKNISSKYADERDRABAEAREKETKALSALARALEALEAKEELERT 1505
Qy 691 KAIEKGE---AVSQEAEILRHRFRLEEAKSRDLQAEVQALKEKIHLMNK-----EDQLSQL 743
Db 1506 NKMLKAEMEDLVSSKDDVGNVHVELEKSKRALETQMEEMKQTQLEEELELQATEDAKURL 1565
Qy 744 QVDYSV-----LQRFMEETKKNMGREVNLTKLELSKRYSRAL 785
Db 1566 EVNMQALKGQFERDLQARDEQNEEKRRQLRQLHEYETELE-DERKQRAL 1614

RESULT 8

MYHB_MOUSE STANDARD; PRT; 1972 AA.
AC O08638; O08639; Q62462; Q64195;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, smooth muscle isoform (SMMHC).
GN MYH11.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Uterus;
RX MEDLINE=97242182; PubMed=9125171;
RA Hasegawa K., Arakawa E., Oda S., Matsuda Y.;

RT "Molecular cloning and expression of murine smooth muscle myosin heavy
RT chains.",
RL Biochem. Biophys. Res. Commun. 232:313-316(1997).
RN [2]
RP SEQUENCE OF 1-368 FROM N.A.
RC TISSUE=Uterus;
RX MEDLINE=95008063; PubMed=7923625;
RA Miano J.M., Cserjesi P., Ligon K.L., Periasamy M., Olson E.N.;
RT "Smooth muscle myosin heavy chain exclusively marks the smooth muscle
RT lineage during mouse embryogenesis.";
RL Circ. Res. 75:803-812(1994).
RN [3]
RP SEQUENCE OF 1-126 FROM N.A.
RC TISSUE=Uterus;
RX MEDLINE=96172919; PubMed=8593698;
RA Suzuki T., Kim H.S., Kurabayashi M., Hamada H., Fujii H., Aikawa M.,
RA Watanabe M., Watanabe N., Sakomura Y., Yazaki Y., Nagai R.;
RT "Preferential differentiation of P19 mouse embryonal carcinoma cells
RT into smooth muscle cells. Use of retinoic acid and antisense against
RT the central nervous system-specific POU transcription factor Brn-2.";
RL Circ. Res. 78:395-404(1996).
CC -!- FUNCTION: Muscle contraction.
CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
CC and 2 regulatory light chain subunits (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=O08638-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O08638-2; Sequence=VSP_003346;
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
CC meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be
CC split further into 2 globular subfragments (S1) and 1 rod-shaped
CC subfragment (S2).
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D85923; BAA19690.1; -.
DR EMBL; D85924; BAA19691.1; -.
DR EMBL; L25860; AAA67552.1; -.
DR EMBL; S81516; AAB36168.1; -.
DR PIR; I52863; I52863.
DR PIR; JC5420; JC5420.
DR PIR; JC5421; JC5421.
DR HSSP; P10587; 1BR2.
DR MGD; MGI:102643; Myh11.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;

FT	DOMAIN	837	1925	COILED COIL (POTENTIAL).
FT	NP_BIND	174	181	ATP (POTENTIAL).
FT	DOMAIN	654	676	ACTIN-BINDING.
SQ	SEQUENCE	1959	AA; 226502 MW; A75C86086FD3A1A1 CRC64;	
Query Match				
Best Local Similarity 7.4%; Score 447.5; DB 1; Length 1959;				
Matches 224; Conservative 192; Mismatches 344; Indels 247; Gaps 41;				
QY	40	KANRKEEDVMASGTIKRHLKPSGESEKTKKSVLSKEDLIQLLSIMEGELQAREDIVHM	99	
Db	839	QVSRQEEEMMA-----KEE--ELIKVKEKQLAAENRLSEM	871	
QY	100	LRTEKTKPEVLBAHYGSAEPKVLRLVLRDAILAQEKSIGEDVYEKPISELDRLEEKQKE	159	
Db	872	-----ETFQAQL-MAEKMQLOEQLOAEAEALCAE---ABEIRARLTAKKQOELEEEICH	919	
QY	160	TYRRMLEQLLLAEKCHRRVTVYELENEKHKHHTDYMNKSDDFTNLLQEQRERLAKLLLEQEK	219	
Db	920	LEARVEEE--EERCQ-----HLQAEKK-----MQQNIQBLEEQLEEEES	957	
QY	220	YQARKEKENA---KRLNKLRLDELVKLKSFAIMLVDERQMHIQLGLQSKVQDLTQKLRE	276	
Db	958	ARQKLQLEKVTTEAKLKLBEEDVIVLEDQNLKLAKEKK-----LLEDRMSEFTTNLT	1010	
QY	277	EEKLKAVT-----YKSKEDRQKLLKLEVDFFHKAS-----	307	
Db	1011	EEKSKSLAKLKNKHEAMITDLEERLRREEQRQOELEKTRRKLEGDSSDLHDQIAELQAO	1070	
QY	308	-----RFSQHEHEEMNAKLANQESHNRQLRLKVLGLSORIEELETNKSILQKAAEEELQ	361	
Db	1071	IAELKIQLSKKEEELQALARVEEAAQKNMAL---KKIRELESQITELQ---EDLESE	1123	
QY	362	REKIAKGECGNSSLMAEVESLRKV---LEMEGKDDEI-TKTEAQCRELKKKLQEE-BHH	416	
Db	1124	RASRNKAEKOKRDLGEELEALKTELEDITLDSTAAQOELRSKREQEVTVLKKTLEDEAKTH	1183	
QY	417	SKELRLEVEKLQKRMSELEKLEAEAFSRKSECTQLHLNLEKEKNLTK----DLLNELEV-	471	
Db	1184	EAQIQ---EMRQKHSQAIEELAEQLEQTK-----RVKANLEKAKQALESERAELSNEVK	1236	
QY	472	-----VKSRVKELECSERULEKAELSRLKDDLTCLKSFTVMLVDERKNNMEKI	518	
Db	1237	LOGKGAEHKRXKVDQAQLQELQVKFTGERVKTELAERNVKLV-----ELDNVTGLL	1289	
QY	519	KQEERKVDGLNKNFKVEQKVMVDVTEKLTIEESKLLKLLKSEMEEEKYSILTKERDELMGKL	578	
Db	1290	NQSDSKSIKLAQFSALESQLODQTQELLQOEBTRLKLSFSTKLQTE-----DEKNALKBQL	1345	
QY	579	RSEERSCELSCSVLL-----KKRLD---GIEEVBREINRGRSCKGSEFTCPEDNKI	628	
Db	1346	EEEEAKRNLEKQISVLQQAQAVEARKKMDGGLGCLTAIEEAKKLLQKDLESLTQRYEEKI	1405	
QY	629	RELTLEIERLKKRLQO-LEVVEGD-----LMKTEDEYDQL-----EQKFRTEQ	670	
Db	1406	AAVD-KLEKTKTRLQOELDDIAVDLDHQRTVSNLEKKQKKFDQLLAEKNISAKYAEER	1464	
QY	671	DKANFLSQOLEEIKHQMAKHKAIEKGEAVSQEAELRH---RFR-----	710	
Db	1465	DRAEAAREKETKALSIA--RALE--EAIEQKAELERNVKQFRTEMEDLMSSKODVGKSV	1520	
QY	711	--LEEAKSRDLQAEVQALKEKIHLMNKK-----EDQLSQLOVDYSVLQORF-----ME	756	
Db	1521	HELEKAK-RALEQQQVEEMKTQLEEELEDELOATEDAKLRLLEVNVQQAQAFDRDLLGRDEQ	1579	
QY	757	EETKNKMGREVLNLTKELELSKRYSRALRPSGNGRRMVDVPVASTGVQTEAVCGDAABEE	816	
Db	1580	NEEKRLQIRQVREMEVELE-DEKQRSIAVAARKLELDLKDLESHIDTANKNRDEAIK	1638	
QY	817	ETPAVFIR-KSFQEE-----NHIMSNLRQVGLK-KPMERSSVLDRYPPAANELTMRK	866	
Db	1639	HVRKLAQAMQMDYMLEDTRTSREEILAQAKENKLLKSMEAEIMQLQEELAAAEAKRQ	1698	

QY	867	SWIPWMRKRENGPSTPQEKGRPNQAGHPGELVLAPKQOQPLHIRV	913	
Db	1699	A-----QQRDELADEIANSSGKGALAMEEKKRLEARI	1731	
RESULT 10				
ID	REST_HUMAN	STANDARD;	PRT;	1427 AA.
AC	P30622;			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-APR-1993 (Rel. 25, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Restin (Cytoplasmic linker protein-170 alpha-2) (CLIP-170) (Reed-Sternberg intermediate filament associated protein).			
GN	RSN.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Peripheral blood monocytes;			
RX	MEDLINE=92289675; PubMed=1600942;			
RA	Bilbe G., Delabie J., Brueggen J., Richener H., Asselbergs F.A.M., Cerletti N., Sorg C., Odink K., Tarsay L., Wiesendanger W., de Wolf-Peeters C., Shipman R.;			
RA	"Restin: a novel intermediate filament-associated protein highly expressed in the Reed-Sternberg cells of Hodgkin's disease.";			
RT	EMBO J. 11:2103-2113(1992).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=92405160; PubMed=1356075;			
RX	Pierre P., Scheel J., Rickard J.E., Kreis T.E.;			
RA	"CLIP-170 links endocytic vesicles to microtubules.";			
RL	Cell 70:887-900(1992).			
CC	-!- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.			
CC	-!- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE CYTOSKELETON.			
CC	-!- ALTERNATIVE PRODUCTS:			
CC	Event=Alternative splicing; Named isoforms=2;			
CC	Name=Long;			
CC	IsoId=P30622-1; Sequence=Displayed;			
CC	Name=Short;			
CC	IsoId=P30622-2; Sequence=VSP_000765;			
CC	-!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS OF HODGKIN'S DISEASE.			
CC	-!- SIMILARITY: Contains 2 CAP-Gly domains.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X64838; CAA46050.1; -.			
DR	EMBL; M97501; AAA35693.1; -.			
DR	PIR; S22695; S22695.			
DR	Genew; HGNC:10461; RSN.			
DR	MIM; 179838; -.			
DR	GO; GO:0005768; C:endosome; TAS.			
DR	GO; GO:0005882; C:intermediate filament; TAS.			
DR	GO; GO:0015630; C:microtubule cytoskeleton; TAS.			
DR	GO; GO:0008017; F:microtubule binding; TAS.			
DR	GO; GO:0006899; P:nonspecific vesicle transport; TAS.			
DR	InterPro; IPR000938; CAP-Gly.			
DR	InterPro; IPR001878; Znf_CCHC.			
DR	Pfam; PF01302; CAP_GLY_2.			
DR	SMART; SM00343; Znf_C2HC; 1.			
DR	PROSITE; PS00845; CAP_GLY_1; 2.			
DR	PROSITE; PS50245; CAP_GLY_2; 2.			

KW		Cytoskeleton; Microtubule; Coiled coil; Repeat; Alternative splicing.	
FT	DOMAIN	78 120	CAP-GLY 1.
FT	DOMAIN	143 204	SER-RICH.
FT	DOMAIN	232 274	CAP-GLY 2.
FT	DOMAIN	304 331	SER-RICH.
FT	DOMAIN	350 1342	COILED COIL (POTENTIAL).
FT	DOMAIN	1408 1421	CCHC-BOX.
FT	VARSPLIC	457 491	Missing (in isoform Short).
FT	CONFLICT	1069 1069	/FTid=VSP 000765.
FT	SEQUENCE	1427 AA; 160989 MW; 0A4F166DD94254E8 CRC64;	D -> E (IN REF. 2).
SQ			
QY		Query Match 7.3%; Score 447; DB 1; Length 1427;	
Db		Best Local Similarity 22.3%; Pred. No. 7.2e-08;	
Matches		196; Conservative 187; Mismatches 329; Indels 166; Gaps 35;	
QY	34	EDAKKNKANRKEEDVMASGT-----IKRHLKPSGESEKTKKSVLSKE--DLIQLL 83	
Db	476	EKTAKDKLQRELEDTRVATVSEKSRIMELEKDLALRVQVAELRRRLESNKPAGVDMSL 535	
QY	84	SIMEGELQAREDIVHMLRTEKTKPEV--LEAHYGSAP--EKVLRLVLRDAILAEQKSIG 139	
Db	536	SLIQ-EISSLOEKLEVTRTDHQR-EITSLKEHFGAREETHQKEIKALY----TATEKLSK 589	
QY	140	EDVYEKPISELDRLLEKQKETYRRMLEQLLLAEKCHRTVYVELENEKHK----- 188	
Db	590	EN--ESLSKLEHANKENS:DVIALWKSLETAIAHQQAAMEELKVSFSGLTETAEFAE 647	
QY	189	-----HTDYMNKSDDFTNLLQEERERLKKLE--QEKAYQARKEKENA-----KRLN 233	
Db	648	LKTQIEKMRLDYQHEIENLQOQDSERAAHAKEMEALRAKLMKVIKEKENSLEAIRSKLD 707	
QY	234	KLRDE-LVKLSFALMLVDERQMHIQLGL-----QSQKVQDLTQKLREEEEK--LK 282	
Db	708	KAEDQHLVEMED-TLNKLQAEIKVKEVLQAKCNEQTKVIDNFTSQLKATEEKLLDLD 766	
QY	283	AVTYKSKEDRQKLLKLEVDPEHKASRFSQEHEEMNAKLANQESHNRQLR---LKLVLGSLQ 339	
Db	767	ALRKASSEGKSEMKKLROQLEAAEKQIKHLEIEKNAESSKASSITRELQGRELKLTLNQE 826	
QY	340	RIEELEETNKSLOKAAEELQELREKIAKG-----ECGN-----SSLMAE 378	
Db	827	NLSEVSQVKETLEK---ELQILKEKFAEASEEAVSVQSRMQETVKNLKHQKEQFNMLSSD 883	
QY	379	VESLRKRVLMEGKDEEITKTEAQCRELKKQLQEE-----KQKLEEE----- 413	
Db	884	LEKLRENLADEAKFREKDEREEQLIKAKEKLENDIAEIMKMSGDNSSQLTKMNDLRLK 943	
QY	414	EHHSKELRLEVEKLRMSLEKLEEAFFS---RSKSPECTQLHLNLEKEKNLTKDILNEL 469	
Db	944	ERDVEELQLKLTKANENASFLQKSIEDMTVKAEQSQQAAKKH---EEEKKELEKLSDL 1000	
QY	470	EVVKSrvKELECSERLE--KAELSLKDDLTKLKSFVTLVDERKNMMEKIKQEEKVDG 527	
Db	1001	E-----KKMETSHNQCELKARYERATSETKTKHEEIL-----QNLQKTLTLDTEDKLKG 1049	
QY	528	LKNFKVEQGVMDVTE--KLIEESKLLKLSKSEMEKBYSLTKERDELMGKLRSEERS 585	
Db	1050	AREE---NSGLLQLEELRQADKAKAAQTAEDAMQIME-QMTKEKTETLASLEDTKQTN 1105	
QY	586	CELSCSVDLLKKRLDGIIEVEREINRGRSCKGSEFTCPEDNKIRELTLEIERLK----KR 641	
Db	1106	AKLQNELDTLKE--NNLKNVE-ELN-----KSKELTVENQKMEEFKEIETLKQAAAQK 1157	
QY	642	LQOLEVVEGLDMKTEDEYDQLEQKFRTEQ---DKANFLSQOLEEIKHQMAKHAIEKGEA 698	
Db	1158	SQQLSALQAEENVKLAELGFSRDEVTSHQKLEEERSVLNNQLLEMKKRESKFIK----DA 1213	
QY	699	VSQEAELRHRFRLEEAKSRDLQAEVQALKEKIHLMNKKEDQLSQLQVDYSVLQRFMEEE 758	
Db	1214	DEEKASLQKSISITSALLTEKDAELEKLRNEVTVLRGENASAKSLHSVVTLESQKVKLE 1273	
QY	759	TKNKNMGREVLNLTKELELSKRYSRALRPSGNRRMVD 796	

Db	1274	LKVKNL-----ELQLENKRQLSSSSGNTDTQAD 1302	
RESULT 11			
MYHB_RABIT			
ID	MYHB_RABIT	STANDARD;	PRT; 1972 AA.
AC	P35748;		
DT	01-JUN-1994 (Rel. 29, Created)		
DT	01-FEB-1996 (Rel. 33, Last sequence update)		
DT	15-MAR-2004 (Rel. 43, Last annotation update)		
DE	Myosin heavy chain, smooth muscle isoform (SMMHC).		
GN	MYH11.		
OS	Oryctolagus cuniculus (Rabbit).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.		
OX	NCBI_TaxID=9986;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=92073350; PubMed=1961735;		
RA	Babij P., Kelly C., Periasamy M.;		
RT	"Characterization of a mammalian smooth muscle myosin heavy-chain		
RT	gene: complete nucleotide and protein coding sequence and analysis of		
RT	the 5' end of the gene.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 88:10676-10680 (1991).		
CC	-!- FUNCTION: Muscle contraction.		
CC	-!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2		
CC	heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)		
CC	and 2 regulatory light chain subunits (MLC-2).		
CC	-!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.		
CC	-!- DOMAIN: The rodlike tail sequence is highly repetitive, showing		
CC	cycles of a 28-residue repeat pattern composed of 4 heptapeptides,		
CC	characteristic for alpha-helical coiled coils.		
CC	-!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light		
CC	meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be		
CC	split further into 2 globular subfragments (S1) and 1 rod-shaped		
CC	subfragment (S2).		
CC	-!- SIMILARITY: Contains 1 myosin-like globular head domain.		
CC	-!- SIMILARITY: Contains 1 IQ domain.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; M77812; AAA31395.1; -.		
DR	HSSP; P10587; 1BR2.		
DR	InterPro; IPR000048; IQ_region.		
DR	InterPro; IPR001609; myosin_head.		
DR	InterPro; IPR004009; Myosin_N.		
DR	InterPro; IPR002928; Myosin_tail.		
DR	InterPro; IPR002017; Spectrin.		
DR	Pfam; PF00612; IQ; 1.		
DR	Pfam; PF00063; myosin_head; 1.		
DR	Pfam; PF02736; Myosin_N; 1.		
DR	Pfam; PF01576; Myosin_tail; 1.		
DR	PRINTS; PR00193; MYOSINHEAVY.		
DR	ProDom; PD000355; myosin_head; 1.		
DR	SMART; SM00015; IQ; 1.		
DR	SMART; SM00242; MYSC; 1.		
DR	PROSITE; PS50096; IQ; 1.		
KW	Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;		
KW	Calmodulin-binding; ATP-binding; Methylation; Multigene family.		
FT	DOMAIN 1 784	MYOSIN HEAD-LIKE.	
FT	DOMAIN 785 807	IQ.	
FT	DOMAIN 844 1934	COILED COIL (POTENTIAL).	
FT	DOMAIN 1935 1972	CARBOXYL-TERMINAL.	
FT	NP_BIND 178 185	ATP (POTENTIAL).	
FT	DOMAIN 661 683	ACTIN-BINDING (BY SIMILARITY).	
FT	DOMAIN 763 777	ACTIN-BINDING (BY SIMILARITY).	

FT	MOD RES	129	129	METHYLATION (TRI-)	(POTENTIAL).
SQ	SEQUENCE	1972	AA;	227318	MW; 2061A224288D6A4C CRC64;
	Query Match			7.3%;	Score 445; DB 1; Length 1972;
	Best Local Similarity			21.5%;	Pred. No. 1.2e-07;
	Matches	248;	Conservative	164;	Mismatches 333; Indels 406; Gaps 44;
QY	40	KANRKEEDVMAS	----	GTIK-RHLKPSGESEKTKKSVLS	-----
Db	846	QVTRQEEEMQAKED	ELQIKERQQKAES	ELQELQKHQTQ	LSEKNLLQEQQAETELYAE
QY	76	-----	KEDLIQLLS	MEGELQARE	DIHMLRTEKTK--PEVLEAHYGS
Db	906	AEEMRVRLAAKQ	EEILHEMEARLE	EEEDRGQQLQ	AEARKKVAQOMLDLEEQLEEEAA
QY	123	LRVLHRDAIIAQE	--KSIGEDVYEKPI	SELDRLEEKQK	ETVRRMLEQLL-----LAEC
Db	966	RQKLQLEKVTAE	AKIKKLEDDILVMD	----	DQNNKLSKE--RKLLEERISDLTTNLAE--
QY	175	HRRTVYELENEKH	KH-TDVMNKSDD	FTNLL	-----QERERLKKLLEQE-----
Db	1018	-----	EEBAKAKNLT	KLKNKHESMI	SELEVRLLKKEESRQLEKLRKMDGEASDLHE
QY	218	-----	KAYQARKEKEN	AKRLNKLRLDEL	VKLKSPALMLVDEROMHIEQL--GLQ
Db	1070	QIADLQAOIAEL	KMQLAKKEELQ	AALARLEDE	--TSQKNNAK
QY	264	SQ-----	KVQDLTQKL	REEBEK	LKA-----VTYKSKEDRQK-----LLKLEVD
Db	1129	SERAAARNKAEK	QRDLGEELEAL	KTELED	TLDTTATQOELRAKREQEVTVLKALDEETR
QY	304	-HKA--SRFSQ	EEHEMNAKLANQ	-----	ESHNRQLRLKVLGLSQRIE
Db	1189	SHEAQVQEMRQ	KHTQVVEELTE	QLEQFKRAKAN	LDKTKQTLEKENADLAGELRVLGQAKQ
QY	343	ELEETNKSLOK	AEELQELREK	IAKGE	CNGNS-----SLMAEVESLRKRVL
Db	1249	EVEHKK--	KLEVQLQELQ	SKCSDGERARA	EALNDKVHKLQNEVESVTGMLSE
QY	393	-DEEITKTEA	QRELKKLQEE	-----	-----EHH
Db	1306	LAKEVASLGS	QLQDTQELLQ	QETTRQKLN	VSTKLRQLEDERNSLQEQLEDEMEAKQNLERH
QY	417	SKELRLEV	-----	EKLQRMSELEK	LEEA
Db	1366	ISTLNIQLSD	SKKLQDF	ASTVESLEEG	KKRFQKEIESITQQYEEKAAAYDKLEKTKNRL
QY	463	KDLLNELEV	VYKSRVKE	LCSESRLK	EALSLKDDLTKLKSF
Db	1426	QOELDDLVD	LDNQRL---	VSNLEKKQK	FDQLLAEENKISSKYADERDRAEAAREKE
QY	523	RK-----	-----	VDGLNKNF	KVEQKVM----DVTEKLI
Db	1483	TKALSLARALE	EALEAKEBEL	RTNKMLKA	EMEDLVSSKDDVGKNVHELEKSKRALETQME
QY	562	EKEYSLTKER	DELM----	GKLRSE	-----EERSCELSCSVD---
Db	1543	EMKTQLELE	DELQATEDAK	LRLEVN	MQALKVQFERDLQARDEQNEEKRRQLQRLHEYE
QY	594	-----	LLKKRLDG	-----	-----IEEVEREIN
Db	1603	TELEDERKQ	RALAAAKKK	LEGDLK	OLELQADS
QY	611	RGRSCKSE	FTCPEDNK	IRELTLE	-----I
Db	1663	DARASRDE	IFATAKEN	EKKAKSLEAD	LMQLQEDLAAAF
QY	650	-----	GDLMKTE	DEYDQLEQ	KFRTEQ-----DKANFL
Db	1723	GRNALQDEK	RRL	LEARIAQLE	EEEEEQGNMEAMSDRVRKATQQAELS
QY	696	GEAVSQEAE	-----	LRHFR	-----LE-----EAKSRDLQAEVQ

Db	1783	NESARQQLERQ	NKELSKSLQ	EMEGAVKSKFK	STIAALEAKIAQLEEQVEAREKQAAAK
QY	725	ALKE----	KIHELM	-----	-----NKEDQLS
Db	1843	ALKQRDKKL	KEMLLQ	VEDERKMAEQY	KEQAEGNAKVQLK
QY	750	LQORFME	EETKNQNMGR	EVNLNLTKE	LELSKRYSRALRPSGNGRRMVDVPVASTGVQTEAV
Db	1903	LQRELD	EATESNEAMGR	EVNALKSKLR	RGNETSFVPTRRSGGRRVIE-----
QY	810	CGDAA	EEETPA	820	
Db	1950	NADG	SEEEVDA	1960	
RESULT 12					
ID	RA50	PYRAB	STANDARD;	PRT;	880 AA.
AC	Q9UZC8;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	DNA double-strand break repair rad50 ATPase.				
GN	RAD50 OR PYRAB12200 OR PAB0812.				
OS	Pyrococcus abyssi.				
OC	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;				
OC	Pyrococcus.				
OX	NCBI_TaxID=29292;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=GE5 / Orsay;				
RX	MEDLINE=22511545; PubMed=12622808;				
RA	Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,				
RA	Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,				
RA	Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;				
RT	"An integrated analysis of the genome of the hyperthermophilic				
RT	archaeon Pyrococcus abyssi.";				
RL	Mol. Microbiol. 47:1495-1512(2003).				
CC	-!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The				
CC	rad50/mre11 complex possesses single-strand endonuclease activity				
CC	and ATP-dependent double-strand-specific exonuclease activity.				
CC	Rad50 provides an ATP-dependent control of mre11 by unwinding				
CC	and/or repositioning DNA ends into the mre11 active site (By				
CC	similarity).				
CC	-!- SUBUNIT: Forms a complex with mre11 (By similarity).				
CC	-!- SIMILARITY: Belongs to the SMC family. RAD50 subfamily.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; AJ248286; CAB50131.1; --				
DR	PIR; F75103; F75103.				
DR	HMAP; MF_00449; --; 1.				
DR	InterPro; IPR003593; AAA ATPase.				
DR	InterPro; IPR003439; ABC transporter.				
DR	InterPro; IPR007517; Rad50_zn_hook.				
DR	InterPro; IPR003395; SMC N.				
DR	Pfam; PF04423; Rad50_zn_hook; 1.				
DR	Pfam; PF02463; SMC_N; 1.				
DR	ProDom; PD000006; ABC transporter; 1.				
DR	SMART; SM00382; AAA; 1.				
KW	DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.				
FT	NP_BIND	30	37	ATP (BY SIMILARITY).	
FT	DOMAIN	144	745	COILED COIL (POTENTIAL).	
SQ	SEQUENCE	880	AA;	103970	MW; FDB177EC7E026479 CRC64;
Query Match					
Best Local Similarity					
7.3%; Score 441; DB 1; Length 880;					
23.7%; Pred. No. 6.9e-08;					

FT DOMAIN 1 790 MYOSIN HEAD-LIKE.
FT DOMAIN 791 820 IQ.
FT DOMAIN 849 1978 RODLIKE TAIL (S2 AND LMM DOMAINS).
FT DOMAIN 849 1978 COILED COIL (POTENTIAL).
FT NP BIND 176 183 ATP.
FT DOMAIN 666 688 ACTIN-BINDING.
FT DOMAIN 767 781 ACTIN-BINDING.
FT MOD RES 1 1 BLOKED.
FT MOD RES 127 127 METHYLATION (TRI-) (POTENTIAL).
FT CONFLICT 127 127 MISSING (IN REF. 3).
FT CONFLICT 204 215 KOTSITQGPSFS -> RTPASLKVHLP (IN REF. 1).
FT STRAND 34 37
FT STRAND 43 47
FT STRAND 56 60
FT STRAND 66 69
FT TURN 72 73
FT STRAND 75 76
FT HELIX 80 82
FT TURN 83 84
FT STRAND 87 87
FT TURN 88 89
FT HELIX 96 108
FT TURN 109 110
FT STRAND 113 115
FT STRAND 120 123
FT TURN 130 131
FT HELIX 134 139
FT TURN 140 142
FT HELIX 145 147
FT HELIX 152 166
FT TURN 167 167
FT STRAND 170 175
FT TURN 178 179
FT HELIX 182 197
FT HELIX 218 223
FT TURN 224 225
FT HELIX 226 234
FT STRAND 235 236
FT TURN 242 243
FT STRAND 244 245
FT STRAND 248 255
FT TURN 257 258
FT STRAND 261 269
FT HELIX 273 276
FT TURN 277 277
FT TURN 281 282
FT STRAND 286 286
FT HELIX 287 295
FT HELIX 298 301
FT TURN 302 305
FT TURN 309 311
FT TURN 313 314
FT TURN 316 317
FT HELIX 327 341
FT TURN 342 342
FT HELIX 345 363
FT STRAND 367 367
FT STRAND 376 376
FT HELIX 381 390
FT TURN 391 391
FT HELIX 394 402
FT HELIX 419 449
FT STRAND 458 464
FT STRAND 474 474
FT HELIX 476 506
FT TURN 507 508
FT HELIX 521 528
FT HELIX 536 543
FT HELIX 551 561
FT TURN 562 564
FT TURN 567 567
FT STRAND 568 570
FT STRAND 579 584

FT TURN 585 586
FT STRAND 587 592
FT TURN 594 595
FT HELIX 596 601
FT HELIX 606 613
FT TURN 614 614
FT HELIX 618 623
FT TURN 624 624
FT HELIX 658 674
FT TURN 675 675
FT STRAND 677 684
FT TURN 692 693
FT HELIX 697 706
FT TURN 707 708
FT HELIX 709 718
FT STRAND 722 725
FT HELIX 726 733
FT HELIX 734 737
FT TURN 738 739
FT HELIX 748 759
FT TURN 760 760
FT HELIX 763 765
FT STRAND 766 768
FT STRAND 772 775
FT TURN 777 778
FT HELIX 779 787
SQ SEQUENCE 1978 AA; 228663 MW; B7B6C923E5273D93 CRC64;

Query Match 7.3%; Score 441; DB 1; Length 1978;
Best Local Similarity 23.3%; Pred. No. 1.6e-07;
Matches 236; Conservative 164; Mismatches 364; Indels 250; Gaps 41;

QY 40 KANRKEEDVMASG-----TIKRHLKPSGSEBKTKKSVEL----- 74
Db 851 QVTRQEEEMQAKDEELQRTKEROQKAEALKEQKHTQLCBEKNLLQEKLOABTELYAE 910
QY 75 -----SKEDLIQLLSIMEGELQAREDVHMLRTEKTK--PEVLEAHYGSAPPEKV 122
Db 911 AEEMRVRLAAKKQOELEELHMEARIEEEEEERSQQLQAEKKKMQQMLDLEEQLSEEEAA 970
QY 123 LRVLHRDAILAQEK-----SIGEDVYEKPISELDRLEEKQKETYRRMLQQLLLAEKC 174
Db 971 RQKLQLEKVTADGKIKKMEDDILIMEDQNNKLTKERKLEERVSDLTNNLAE----- 1022
QY 175 HRRTVYELENEKHKH-TDYMNKSDDFTNNLE-----QERERLKKLLEQE----- 217
Db 1023 -----EERAKNLTCLKNKHESMISELEVRLLKKEEKSROLEKIKRKLEGESSDLHE 1074
QY 218 -----KAYQARKEKENAKRLNKLRLDELVLKLSFALMLVDERQMHIQL--GLQ 263
Db 1075 QIAELQAIHELKAQLAKKEEELQAALARLEDE-TSQNNALKKIRELESHISDLQEDLE 1133
QY 264 SQ-----KVQDLTQKLREEEELKAVTYKSKEDRQKLLKLVDFEHKASRFSQHEEMNA 318
Db 1134 SEKAARNKAEKQKRDLSSELEALKT-----ELED-----TLDTTATQOQLRAKREQEVTVL 1184
QY 319 KLANQE---SHNRQLRLKLVGLSQRIEELLETNKSLOKAEELQELREKIAKGECGNSSL 375
Db 1185 KRALEETRTHEAQVQEMRQKHTQAVEELTEQLEQFKRAKANLDKTKQTTLEK---DNADL 1241
QY 376 MAEVESLRKRVLEMEGKDEEITKTEAQCRELKKLQEEHHHSKEILRLEVEKLO----- 428
Db 1242 ANEIRLSQAKQDVEHKKK---KLEVQLQDLSQKYSQDGERVTRTELNEKVKHKLQIEVENVT 1298
QY 429 -----KRMSELEKLEEFASRSKSECTQLHLNLEKEKNLT KDLLN 467
Db 1299 SLLNEAESKNIKLTKDVATLGSQLODQTQELLQEETROKLNVTTKRLQLEDDEKNSLQEQLD 1358
QY 468 ELEVVKSRVKELECSERLEKABLSLKDDLTCLKSFTVMLVDERKNMMEKIKQEERKVDG 527
Db 1359 E-----EVEAKQNLERHISTLTITQLSDSKKQLQEFATV----ETMEEGKKKLQREIES 1408
QY 528 LKNKF-----KVEQKG-----VMDVTEKLIIESKLLKLKSEM--EEKEY 565

Db 1409 LTQFEEKAASYDKLEKTKNRLQQLDLDLVNDQRLVSNLEKKQKQKFDQMLAEKNI 1468

QY 566 S--LTKERDELMGKLRSSEERSCELSVDLLKKRLDGIIEVEREINRGRSCKGSEFTCP 623

Db 1469 SSKYADERDRAEAAREKETKALSARA--LEEALEAKEELER-TNKMMLKAMEDLVSS 1524

QY 624 EDNKIRELTLEIERLKKRL-QQLEVVEGDLMTKTEDEYDQLEQ-KFRTQDQKANFLSQ--- 678

Db 1525 KDD-VGKNVHELEKSKRTLEQQVEEMKTQLEEDLELQAAEDAKLRLEVNMQAMKSQFER 1583

QY 679 -----QLEEIKHQMAKHKAIEKGEAVSQEALRHRFRLEBAKS-----RDLQAEV-- 723

Db 1584 DLQARDEQNEEKRRQLLKQLHEHETEL---EDERKQALAAAKKLEVDVKDLESQVDS 1640

QY 724 --QALKEKIHELMNKEDQLSLQVDY----SVLQORF---MEEETKNKNMGREVLNLTKE 774

Db 1641 ANKAREEAIKQLRKLQAQMKDYQRDLDDARAAREEIFATARENEKKAKNLEALIQLQED 1700

QY 775 LELSKRYSRALRPSGNGRRMVDVPVASTGVQTEAVCGDAAEEETPAVFIRKSPQEE-NHI 833

Db 1701 LAAAE-----ARKQAD-----LEKE-----EMAEELASANSGRTSIQDEKRRL 1739

QY 834 MSNLRQVGLKKPMERSV---LDRYPPAA-----NELTMRKSWIPWMRKREN 877

Db 1740 EARIAQLEELDEHSNIETMSDRMRKAVQQAQEQNLNELATERA---TAQKNEN 1790

RESULT 14

EEAL_HUMAN STANDARD; PRT; 1411 AA.

AC Q15075; Q14221;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Early endosome antigen 1 (Endosome-associated protein p162).

GN EEAL.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.

RC TISSUE=Cervical carcinoma;

RX MEDLINE=95286647; PubMed=7768953;

RA Mu F.-T., Callaghan J.M., Steele-Mortimer O., Stenmark H.,

RA Parton R.G., Campbell P.L., McCluskey J., Yeo J.-P., Tock E.P.C.;

RA Toh B.-H.;

RA "EEAL, an early endosome-associated protein. EEAL is a conserved

RT alpha-helical peripheral membrane protein flanked by cysteine

RT 'fingers' and contains a calmodulin-binding IQ motif.";

RL J. Biol. Chem. 270:13503-13511(1995).

[2]

RN SEQUENCE FROM N.A.

RP Seelig H.P.;

RA Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.

[3]

RP INTERACTION WITH RAB5A.

RX MEDLINE=98361239; PubMed=9697774;

RA Simonsen A., Lippe R., Christoforidis S., Gaullier J.-M., Brech A.,

RA Callaghan J.M., Toh B.-H., Murphy C., Zerlial M., Stenmark H.;

RT "EEAL links PI(3)K function to Rab5 regulation of endosome fusion.";

RL Nature 394:494-498(1998).

[4]

RN INTERACTION WITH RAB5A AND RAB5B.

RP MEDLINE=99421752; PubMed=10491193;

RX Callaghan J.M., Nixon S., Buccic C., Toh B.-H., Stenmark H.;

RA "Direct interaction of EEAL with Rab5b.";

RL Eur. J. Biochem. 265:361-366(1999).

[5]

RN INTERACTION WITH STX6, AND SUBCELLULAR LOCATION.

RP MEDLINE=99436077; PubMed=10506127;

RA Simonsen A., Gaullier J.-M., D'Arrigo A., Stenmark H.;

RT "The Rab5 effector EEAL interacts directly with syntaxin-6.";

RL J. Biol. Chem. 274:28857-28860(1999).

[6]

RN MUTAGENESIS OF ASP-1352; ASN-1357; 1367-VAL-THR-1368; ARG-1375 AND

RP ARG-1400, HOMODIMERIZATION, AND INTERACTION WITH PHOSPHATIDYLINOSITOL

RP 3-PHOSPHATE.

RX MEDLINE=99322673; PubMed=10394369;

RA Kutateladze T.G., Ogburn K.D., Watson W.T., de Beer T., Emr S.D.,

RA Burd C.G., Overduin M.;

RT "Phosphatidylinositol 3-phosphate recognition by the FYVE domain.";

RL Mol. Cell 3:805-811(1999).

[7]

RN MUTAGENESIS OF TRP-1349; CYS-1358; PHE-1365; ARG-1370; ARG-1371;

RP HIS-1372; HIS-1373; CYS-1374; ARG-1375; CYS-1377; GLY-1378; CYS-1385;

RP ARG-1400 AND CYS-1405, SUBCELLULAR LOCATION, AND INTERACTION WITH

RP PHOSPHATIDYLINOSITOL 3-PHOSPHATE.

RX MEDLINE=20387352; PubMed=10807926;

RA Gaullier J.-M., Roennig E., Gillooly D.J., Stenmark H.;

RT "Interaction of the EEAL FYVE finger with phosphatidylinositol

RT 3-phosphate and early endosomes. Role of conserved residues.";

RL J. Biol. Chem. 275:24595-24600(2000).

[8]

RN INTERACTION WITH RAB22A.

RP MEDLINE=21859373; PubMed=11870209;

RX Kauppi M., Simonsen A., Bremnes B., Vieira A., Callaghan J.M.,

RA Stenmark H., Olkkonen V.M.;

RT "The small GTPase Rab22 interacts with EEAL and controls endosomal

RT membrane trafficking.";

RL J. Cell Sci. 115:899-911(2002).

[9]

RN MUTAGENESIS OF GLU-39; PHE-41; ILE-42; PRO-44; MET-47 AND TYR-60,

RP HOMODIMERIZATION, AND INTERACTION WITH RAB5C.

RX MEDLINE=22499593; PubMed=12493736;

RA Merithew E., Stone C., Eathiraj S., Lambright D.G.;

RT "Determinants of Rab5 interaction with the N terminus of early

RT endosome antigen 1.";

RL J. Biol. Chem. 278:8494-8500(2003).

[10]

RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1289-1411 IN COMPLEX WITH

RP PHOSPHATIDYLINOSITOL 3-PHOSPHATE, AND HOMODIMERIZATION.

RX MEDLINE=21617582; PubMed=11741531;

RA Dumas J.J., Merithew E., Sudharshan E., Rajamani D., Hayes S.,

RA Lawe D., Corvera S., Lambright D.G.;

RT "Multivalent endosome targeting by homodimeric EEAL.";

RL Mol. Cell 8:947-958(2001).

[11]

RN STRUCTURE BY NMR OF 1346-1410 ALONE AND IN COMPLEX WITH

RP PHOSPHATIDYLINOSITOL 3-PHOSPHATE.

RX MEDLINE=21143489; PubMed=11230696;

RA Kutateladze T.G., Overduin M.;

RT "Structural mechanism of endosome docking by the FYVE domain.";

RL Science 291:1793-1796(2001).

CC -1- FUNCTION: Binds phospholipid vesicles containing

CC phosphatidylinositol 3-phosphate and participates in endosomal

CC trafficking.

CC -1- SUBUNIT: Homodimer. Binds STX6. Binds RAB5A, RAB5B, RAB5C and

CC RAB22A that have been activated by GTP-binding.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic and peripheral membrane protein

CC associated with early endosomes.

CC -1- DOMAIN: The FYVE-type zinc finger domain mediates interactions

CC with phosphatidylinositol 3-phosphate.

CC -1- DISEASE: Antibodies against EEAL are found in sera from patients

CC with subacute cutaneous lupus erythematosus and other autoimmune

CC diseases.

CC -1- SIMILARITY: Contains 1 C2H2-type zinc finger.

CC -1- SIMILARITY: Contains 1 FYVE-type zinc finger.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

Db 974 -----| ||: |: QIEALQGEL-----KIA 985

Qy 823 IRKSFQENHIMNLQVGLKPKMERSVLD RYPPAANELTMKSWIPWMRKRENGSPSTP 882

Db 986 VLQKTELENKLQQLTQ-----AAQELAAEKEKISVL---QNNYEKS 1024

Qy 883 QEKGPR-PNQAGHPGELVLAPKQGPLH R VTPDHENSTATLEITSPTSEFFSSTVI 941

Db 1025 QETFKQLQSDFYGRESEL-LATRQ-----DLKSVEEKLSLAQEDLISNRNQI 1070

Qy 942 PTLGNQKPRITIIPSPNVMSQPKSADPTLGP ERAMSP-----VTITTISREKSPGG 994

Db 1071 ---GNQ-----NKLIELKTAKATLEQDSAKKEQQLQERCKALQDIQEKSLKEK 1117

Qy 995 R S A F A D R P A S P I Q I M T V S T S A A P T E I A V S P E S Q E V P M G R T I L K V T P E K Q ----- 1043

Db 1118 ELVNEKSKLAEIEEIKCRQEKBITKLINEELKSHKLESIKEITNLKDAKQLLIQKLELQG 1177

Qy 1044 ---TVPAPVRKYNSNANIITTEDNKIHIH LGSQFKRSPGPAAEGVSPVITVRPVNVTAEK 1100

Db 1178 KADSLKAAVEQEKRNQILKQVKKKEEELKKEFIEKEAKLHSEIK-----EK 1225

Qy 1101 EVSTGTVLRSRPNHLSSRPGASKVTSTITIPVTTSSTRTQSVSGDGSQRPTTRIP 1160

Db 1226 EVGM-----KKH---EENEA K L T M Q I T ---ALNENLGTVKKEWQ---SSQRRVSELEK 1269

Qy 1161 MSKGMKAGKPVVAAS-----GAGNLTKFPRAETQSMKIELKKSAASS 1203

Db 1270 QTDDL RGEI A V L E A T V Q N N Q D R A L L E R C L K G E G E I E K L Q T K V -----L E L O R K L D N T 1323

Qy 1204 T A S L 1207

Db 1324 T A A V 1327

RESULT 15

MYH3_RAT

ID MYH3 RAT STANDARD; PRT; 1940 AA.

AC P12847;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Myosin heavy chain, fast skeletal muscle, embryonic.

GN MYH3.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=8706098; PubMed=3783701;

RA Strehler E.E., Strehler-Page M.-A., Perriard J.C., Periasamy M.,

RA Nadal-Ginard B.;

RT "Complete nucleotide and encoded amino acid sequence of a mammalian

RT myosin heavy chain gene. Evidence against intron-dependent evolution

RL of the rod.";

RL J. Mol. Biol. 190:291-317(1986).

CC -!- FUNCTION: Muscle contraction.

CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2

CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)

CC and 2 regulatory light chain subunits (MLC-2).

CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.

CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing

CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,

CC characteristic for alpha-helical coiled coils.

CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light

CC meromyosin (LMW) and 1 heavy meromyosin (HMM). It can later be

CC split further into 2 globular subfragments (S1) and 1 rod-shaped

CC subfragment (S2).

CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.

CC -!- SIMILARITY: Contains 1 IQ domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X04267; CAA27817.1; -.

DR PIR; A24922; A24922.

DR HSSP; P13538; 2MYS.

DR InterPro; IPR000048; IQ region.

DR InterPro; IPR001609; myosin head.

DR InterPro; IPR004009; Myosin_N.

DR InterPro; IPR002928; Myosin_tail.

DR InterPro; IPR002017; Spectrin.

DR Pfam; PF00612; IQ; 2.

DR Pfam; PF00063; myosin head; 1.

DR Pfam; PF02736; Myosin_N; 1.

DR Pfam; PF01576; Myosin tail; 1.

DR PRINTS; PR00193; MYOSINHEAVY.

DR ProDom; PD000355; myosin_head; 1.

DR SMART; SM00015; IQ; 1.

DR SMART; SM00242; MYSC; 1.

DR PROSITE; PS50096; IQ; 1.

DR MYOSIN; Muscle protein; Coiled coil; Thick filament; Actin-binding;

KW Calmodulin-binding; ATP-binding; Methylation; Multigene family.

KW

FT DOMAIN 1 781 MYOSIN HEAD-LIKE.

FT DOMAIN 782 811 IQ.

FT DOMAIN 840 1933 COILED COIL (POTENTIAL).

FT NP_BIND 179 186 ATP (POTENTIAL).

FT DOMAIN 656 678 ACTIN-BINDING.

FT DOMAIN 758 772 ACTIN-BINDING.

FT MOD_RES 130 130 METHYLATION (TRI-) (POTENTIAL).

SQ SEQUENCE 1940 AA; 223857 MW; B5D546A596E5A696 CRC64;

Query Match 7.2%; Score 439; DB 1; Length 1940;

Best Local Similarity 21.6%; Pred. No. 1.8e-07;

Matches 223; Conservative 181; Mismatches 338; Indels 292; Gaps 41;

Qy 33 SEDAKKNKANRKEE-----DV MASGTIKRHLKPSGESEKTKKSVELSKEDL-IQLLSIM 86

Db 843 SAETEKEMATMKEEFQKTKDELAKSEAKRK-----ELEELVTLVQ-EKNDLQVQVAES 896

Qy 87 EGELQAREDVIHMLRT---E KTKPEVLEAHVCSAEPEKVLRLHRDAILAQEKSIGEDV 142

Db 897 ENLLDAEERCDQLIAKAFQLEAKIKEVTE-----RAEDEEEI-----NAELTAKKRKLEDE 947

Qy 143 YEKPISELDRLE-----EKQK-----ETYRRMLEQLLLAEKCHRRRT 178

Db 948 CSELKKDIDDLLELTAKVEKEKHATENKVKNLTEELAGLDETI AKLTREKKALQEAHQQT 1007

Qy 179 VYELENEKHKHTDYMNKSDDFTNLLEQERERLKKLEQEKAYQARKEKENAKRLNKL R-- 236

Db 1008 LDDLQAE E-----DKVNSLSKLKSKLEQQVDDLESSLEQEKKL RVDLERNKRKLEGLKLA 1063

Qy 237 -----DELVKLKSFALMLVDERQWHIEQLGLQSQ-KVQDLTQKLREEEKL 281

Db 1064 QESILDLENDKQQLDERLKKDFEYSQLQSKVEDEQTL SLQLQKKIKELQARIEELEEEI 1123

Qy 282 -----KAVTYKSKED-----RQKLLKLEVDFFE--- 303

Db 1124 EAERATRAKTEKQSDYARELEELSERLEEAGVTSTQIELNKKREAEFLKRRDLEEAT 1183

Qy 304 --HKA--SRFSQEHHEMNAKLANQ-----ESHNRQLRLKLVGLSORIEELEETN 348

Db 1184 LQHEATVATLRKKHADSAAEAEQIDNLQRVKQKLEKEKSEFKLEIDDLSSSVESVSKS K 1243

Qy 349 KSLQK---AEELQELREK-----IAKGECSNSSLMAEVESLRKRVLEMEGKDEEIT 397

Db 1244 ANLEKICRTLEDQLSEARGKNEETQ RSLSELTTQKSRLQTEAGELSRQLEEKESIVS QLS 1303

Qy 398 KTE---AQCRELKKKLQEE-----EHSKELRLEVEKLQKRMSEL----- 434

Db	1304	RSKQAF	QOIEEL	KRQLBEEN	KAKNALAHALQSSRHDCDLLREQYEEEGKAEQALQS	1363
Qy	435					
Db	1364	KANSEVAQWRTKYETDAIQORTEEELEBAKKLAQRLQDSBEEQVEAVNAKASLEKTKQRLQ	1423			
Qy	460	NLT	KDLLNELE	VVKS	RVKELE-----CSESRL-----KAELS	SLKDDLT 498
Db	1424	GEVEDLMVDVERANS	LAAALDKKQ	RNFDKVLAEWKTKCESQAELEAALKESRSLSTELF	1483	
Qy	499	KLKSF	TVMLVDERKNNMEKIKQEEERKVDGINKNFKVEQGVMDVTEKLIIEESK	---	KLK 555	
Db	1484	KLKN	----	AYEEALDQLETVKRE-----NKNLEQE-----IADLTEQ	IAENGKSIHELEK 1529	
Qy	556	LKSEMEE	KYSLTKERDELMGKLRS	EEERSCSVDLLKKRLDGIIEEVEREINRGRSC	615	
Db	1530	SRKQMELEKADIQMALE	EAAALEHEEAK-----ILRIQLE-LTQVKSEI	-----	1573	
Qy	616	KGSEFT	CPEDNKIRELTLEIERLKKRLQO-LEVVEGDL	---	MKTEDEYDQLEQKFRTEQD 671	
Db	1574	-----	DRKIAEKDEEIEQ	LRNYQRTVETMQGALDAEVRSRNEAIRLKKMEGDLN	1624	
Qy	672	K-----	ANFLSQQLE	EIKHQMA-----KHKAIEKGEAVSQAEALHRFRLEEEAKSRDL	719	
Db	1625	EIEIQLSHAN	--RQAAETIKHLRSVQ	GQLKDTQLHLDDALRGQEDLKEQLAIVERRANLL	1682	
Qy	720	QAEVQALKEKIH	ELMKNKEDQLSOLQVDYSVLQORFMEEETK	----	NKNMGREVLNLTKE 775	
Db	1683	QAEVEELRATLEQ	TERARKLAEQELLDSNERNVOLLHTQNTSLIHTKKKLETDLTQLQSEV	1742		
Qy	776	ELSKRYSRALRPSG	NGRRMVDVPVASTGVQTEAVCGDAEEETPAVF	-----	IRKSFOE 829	
Db	1743	EDASRDAR	-----NAEEKAKKAITDAAMMAEEL	---	KKEQDTSAHLERMKNLEQTVKD 1793	
Qy	830	ENHIMSNLRQVGLK	843			
Db	1794	LQHR	LDEAEQLALK	1807		

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2004, 13:07:18 ; Search time 152 Seconds
(without alignments)
2515.843 Million cell updates/sec

Title: US-10-788-793-2
Perfect score: 6082
Sequence: 1 MRSRNQGGESSNGHVSCP...KIELKKSAASTASLGGGKG 1212

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archheap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6082	100.0	1212	11 Q8K4T4	Q8k4t4 rattus norv
2	5696.5	93.7	1213	4 Q7Z7B0	Q7z7b0 homo sapien
3	5456.5	89.7	1177	4 Q86TC3	Q86tc3 homo sapien
4	5349	87.9	1140	4 Q9ULE5	Q9ule5 homo sapien
5	4838	79.5	965	11 Q8JZS5	Q8jzs5 rattus norv
6	4541	74.7	965	4 Q96SK6	Q96sk6 homo sapien
7	3051	50.2	653	4 Q9NVI8	Q9nvi8 homo sapien
8	2522.5	41.5	610	11 Q9CS72	Q9cs72 mus musculu
9	2406.5	39.6	512	4 Q8N8B9	Q8n8b9 homo sapien
10	1926	31.7	893	4 Q8N6Z0	Q8n6z0 homo sapien
11	1828	30.1	764	4 Q8IUM3	Q8ium3 homo sapien
12	1645.5	27.1	752	4 Q13597	Q13597 homo sapien
13	893	14.7	1026	4 Q86V48	Q86v48 homo sapien
14	891	14.6	1046	4 Q8TEH1	Q8tehl homo sapien
15	839	13.8	1067	11 Q99NG3	Q99ng3 mus musculu
16	833	13.7	1068	11 Q8R4U7	Q8r4u7 mus musculu

17	829	13.6	1051	11 Q9ESV1	Q9esv1 rattus norv
18	532	8.7	201	11 Q9CSL6	Q9csl6 mus musculu
19	497.5	8.2	251	11 Q8BQW1	Q8bqw1 mus musculu
20	475.5	7.8	1305	10 Q9FJ35	Q9fj35 arabidopsis
21	474.5	7.8	2139	5 Q07569	Q07569 entamoeba h
22	474	7.8	2363	4 Q99968	Q99968 homo sapien
23	473	7.8	2055	5 Q8T5C7	Q8t5c7 plasmodium
24	473	7.8	2055	5 Q8IHP3	Q8ihp3 plasmodium
25	471.5	7.8	2007	13 Q02015	Q02015 gallus gall
26	470.5	7.7	1596	5 Q8IJ44	Q8ij44 plasmodium
27	464.5	7.6	1790	3 Q07380	Q07380 saccharomyc
28	464	7.6	1229	5 Q9NJ22	Q9nj22 aequipecten
29	464	7.6	1243	5 Q9NJ21	Q9nj21 aequipecten
30	464	7.6	1253	5 Q9NJ20	Q9nj20 aequipecten
31	464	7.6	1951	5 Q17042	Q17042 aequipecten
32	463.5	7.6	1205	5 P92021	P92021 caenorhabdi
33	463.5	7.6	2003	5 Q19658	Q19658 caenorhabdi
34	463	7.6	1960	11 Q8VDD5	Q8vdd5 mus musculu
35	462.5	7.6	1909	5 Q25893	Q25893 plasmodium
36	461.5	7.6	1999	11 Q63731	Q63731 rattus norv
37	459.5	7.6	2003	5 Q22869	Q22869 caenorhabdi
38	457	7.5	1219	5 Q9NJ23	Q9nj23 aequipecten
39	456	7.5	1972	11 Q8R384	Q8r384 mus musculu
40	455.5	7.5	2328	5 Q9VY43	Q9vy43 drosophila
41	449	7.4	1992	13 Q04834	Q04834 xenopus lae
42	448.5	7.4	1941	5 Q26079	Q26079 placopecten
43	448.5	7.4	1964	13 Q93522	Q93522 xenopus lae
44	447.5	7.4	2360	5 Q8IR55	Q8ir55 drosophila
45	445.5	7.3	1950	5 Q26080	Q26080 placopecten

ALIGNMENTS

RESULT 1

Q8K4T4
ID Q8K4T4 PRELIMINARY; PRT; 1212 AA.
AC Q8K4T4;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Filamin-interacting protein L-FILIP.
GN L-FILIP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar;
RX MEDLINE=22100428; PubMed=12055638;
RA Nagano T., Yoneda T., Hatanaka Y., Kubota C., Murakami F., Sato M.;
RT "Filamin A-interacting protein (FILIP) regulates cortical cell
migration out of the ventricular zone.";
RL Nat. Cell Biol. 4:495-501(2002).
DR EMBL; AB055759; BAC00851.1; -.
DR InterPro; IPR000533; Tropomyosin.
DR PRINTS; PR00194; TROPOMYOSIN.
SQ SEQUENCE 1212 AA; 137752 MW; A1D5B9CSAF7D4F80 CRC64;

Query Match 100.0%; Score 6082; DB 11; Length 1212;
Best Local Similarity 100.0%; Pred. No. 9.2e-232;
Matches 1212; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MRSRNQGGESSNGHVSCP	KSSIISSDGGKGPSEDAKKNKANRKEEDVMASGTIKRHLKP	60
Db	1	MRSRNQGGESSNGHVSCP	KSSIISSDGGKGPSEDAKKNKANRKEEDVMASGTIKRHLKP	60
Qy	61	SGESEKTKKSVELSKEDLIQLLSIMEGELQAREDVIHMLRTEKTKPEVLEAHYGS	AEP	120
Db	61	SGESEKTKKSVELSKEDLIQLLSIMEGELQAREDVIHMLRTEKTKPEVLEAHYGS	AEP	120
Qy	121	KVLRVLHRDAILAQEKSIGEDVYEKPISELDRLEEKQKETYRRML	EQLLLA	180
			EQLLLA	180

Db 121 KVLRLVLRDAIIAQEKSIGEDVVEKPISELDRLEEKQETVRRMLEQLLLAEKCHRRTV 180
QY ELENEXKHHTDYMNKSDDDFTNLLEQERERLKKLLEQEKAYQARKEKENAKRLNKLRLDELV 240
Db 181 ELENEXKHHTDYMNKSDDDFTNLLEQERERLKKLLEQEKAYQARKEKENAKRLNKLRLDELV 240
QY 241 KLKSFALMLVDERQMHIEQLGLQSQKVQDLTQKLREEBEKKAVTYKSKEDRQKLLKLEV 300
Db 241 KLKSFALMLVDERQMHIEQLGLQSQKVQDLTQKLREEBEKKAVTYKSKEDRQKLLKLEV 300
QY 301 DFEHKASRFSQEHHEEMNAKLANQESHNRQLRLKLVGLSQRTEEBEETNKSQKAEELQE 360
Db 301 DFEHKASRFSQEHHEEMNAKLANQESHNRQLRLKLVGLSQRTEEBEETNKSQKAEELQE 360
QY 361 LREKIAKGECSNSSLMAEVESLRKRVLMEGKDEEITKTEAQCRELKKKLQEEHHHSKEL 420
Db 361 LREKIAKGECSNSSLMAEVESLRKRVLMEGKDEEITKTEAQCRELKKKLQEEHHHSKEL 420
QY 421 RLEVEKLOKRMSELEKLEEAFFSRSKSECTQLHLNLEKEKNLT KDLLNELEVVKSRVKELE 480
Db 421 RLEVEKLOKRMSELEKLEEAFFSRSKSECTQLHLNLEKEKNLT KDLLNELEVVKSRVKELE 480
QY 481 CSERLEKAELSLKODLTCLKSFTVMLVDERKNMMEKIKQEBERKVDGLNKNFKVEQGKVM 540
Db 481 CSERLEKAELSLKODLTCLKSFTVMLVDERKNMMEKIKQEBERKVDGLNKNFKVEQGKVM 540
QY 541 DVTEKLIBSKKLLKLKSEMEEKEYSLTKERDELMGKLRESEERSCELSCSVDLLKKRLD 600
Db 541 DVTEKLIBSKKLLKLKSEMEEKEYSLTKERDELMGKLRESEERSCELSCSVDLLKKRLD 600
QY 601 GIEEVEREINRGRCKGSEFTCPEDNKIRELTLEIERLKKRLQQLVEVVEGDLMKTEYD 660
Db 601 GIEEVEREINRGRCKGSEFTCPEDNKIRELTLEIERLKKRLQQLVEVVEGDLMKTEYD 660
QY 661 QLEQKFRTEQDKANFLSQLEEBEIKHQMAKHAIEKGEAVSQEAELRHRFRLEEAKSRLDQ 720
Db 661 QLEQKFRTEQDKANFLSQLEEBEIKHQMAKHAIEKGEAVSQEAELRHRFRLEEAKSRLDQ 720
QY 721 AEVQALKEKIHLMNKEDQLSQLQVDYSVLQORFMEEEETKNKMGREVLNLTKELELSKR 780
Db 721 AEVQALKEKIHLMNKEDQLSQLQVDYSVLQORFMEEEETKNKMGREVLNLTKELELSKR 780
QY 781 YSRALRPSGNRRMVDVPVASTGVQTEAVCGDAEEETPAVFIRKSFQENHIMSNLRQV 840
Db 781 YSRALRPSGNRRMVDVPVASTGVQTEAVCGDAEEETPAVFIRKSFQENHIMSNLRQV 840
QY 841 GLKKPMERSSVLD RYPPAANELTMRKSWTPWMRKRENGPSTPQEKGPRNQGAGHPGELV 900
Db 841 GLKKPMERSSVLD RYPPAANELTMRKSWTPWMRKRENGPSTPQEKGPRNQGAGHPGELV 900
QY 901 LAPKQGQPLHIRVTPDHENSTATLEITSPTSEEFFSSTTVIPTLGNOKPRITIIIPSPNVM 960
Db 901 LAPKQGQPLHIRVTPDHENSTATLEITSPTSEEFFSSTTVIPTLGNOKPRITIIIPSPNVM 960
QY 961 SQPKSADPTLGPERAMSPVTITISREKSPGGRSAFADRPASPIQIMTVSTSAAPTEI 1020
Db 961 SQPKSADPTLGPERAMSPVTITISREKSPGGRSAFADRPASPIQIMTVSTSAAPTEI 1020
QY 1021 AVSPESQSVPMGRTILKVTPEKQTVPAVVRKYNSNANITTEDNKIHLGSGQFKRSPGP 1080
Db 1021 AVSPESQSVPMGRTILKVTPEKQTVPAVVRKYNSNANITTEDNKIHLGSGQFKRSPGP 1080
QY 1081 AAEVSPVITVRPNVTAEKEVSTGTVLRSPRNHLSSRPGASKVTSTITPTVTTSTRG 1140
Db 1081 AAEVSPVITVRPNVTAEKEVSTGTVLRSPRNHLSSRPGASKVTSTITPTVTTSTRG 1140
QY 1141 TQSVSGDGSSQRPPTPRI PMSKGMKAGKGPVVAASGAGNLTKFQPRAEQTQSMKIELKSA 1200
Db 1141 TQSVSGDGSSQRPPTPRI PMSKGMKAGKGPVVAASGAGNLTKFQPRAEQTQSMKIELKSA 1200
QY 1201 ASSTASLGCGKG 1212
|||||

Db 1201 ASSTASLGCGKG 1212
RESULT 2
Q7Z7B0
ID Q7Z7B0 PRELIMINARY; PRT; 1213 AA.
AC Q7Z7B0;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE L-FILIP.
GN L-FILIP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Nagano T., Sato M.;
RT "Human orthologue of L-FILIP.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB086011; BAC77067.1; --
SQ SEQUENCE 1213 AA; 138108 MW; EE4329B03516E6AF CRC64;

Query Match 93.7%; Score 5696.5; DB 4; Length 1213;
Best Local Similarity 93.5%; Pred. No. 1.4e-216;
Matches 1134; Conservative 37; Mismatches 41; Indels 1; Gaps 1;

QY 1 MRSRNOGGESSNGHVSCPSSIISSDGGKGPSEDA-KKNKANRKEEDVMASGTIKRHLK 59
Db 1 MRSRNOGGEASDGHISCPKPSIIIGNAGEKSLSEDAKKKKKSNRKEDDVMASGTIVKRLK 60
QY 60 PSGESEKTKKSVLSKEDLIQLLSIMEGELQAREDIVHMLRTEKTKPEVLEAHYGSABP 119
Db 61 TSCECERKTKSLSKEDLIQLLSIMEGELQAREDIVHMLKTEKTKPEVLEAHYGSABP 120
QY 120 EKVLRLVLRDAIIAQEKSIGEDVVEKPISELDRLEEKQETVRRMLEQLLLAEKCHRRTV 179
Db 121 EKVLRLVLRDAIIAQEKSIGEDVVEKPISELDRLEEKQETVRRMLEQLLLAEKCHRRTV 180
QY 180 YELENEXKHHTDYMNKSDDDFTNLLEQERERLKKLLEQEKAYQARKEKENAKRLNKLRLDEL 239
Db 181 YELENEXKHHTDYMNKSDDDFTNLLEQERERLKKLLEQEKAYQARKEKENAKRLNKLRLDEL 240
QY 240 VKLKSFALMLVDERQMHIEQLGLQSQKVQDLTQKLREEBEKKAVTYKSKEDRQKLLKLE 299
Db 241 VKLKSFALMLVDERQMHIEQLGLQSQKVQDLTQKLREEBEKKAVTYKSKEDRQKLLKLE 300
QY 300 VDFEHKASRFSQEHHEEMNAKLANQESHNRQLRLKLVGLSQRTEEBEETNKSQKAEELQ 359
Db 301 VDFEHKASRFSQEHHEEMNAKLANQESHNRQLRLKLVGLTQRIEELEETNKNLQKAEELQ 360
QY 360 ELREKIAKGECSNSSLMAEVESLRKRVLMEGKDEEITKTEAQCRELKKKLQEEHHHSKE 419
Db 361 ELRDKIAKGECSNSSLMAEVENLRKRVLMEGKDEEITKTESQCRELKKKLQEEHHHSKE 420
QY 420 LRLEVEKLOKRMSELEKLEEAFFSRSKSECTQLHLNLEKEKNLT KDLLNELEVVKSRVKELE 479
Db 421 LRLEVEKLOKRMSELEKLEEAFFSRSKSECTQLHLNLEKEKNLT KDLLNELEVVKSRVKELE 480
QY 480 ECSERLEKAELSLKDDLTCLKSFTVMLVDERKNMMEKIKQEBERKVDGLNKNFKVEQKVV 539
Db 481 ECSERLEKAELSLKDDLTCLKSFTVMLVDERKNMMEKIKQEBERKVDGLNKNFKVEQKVV 540
QY 540 MDVTEKLIBSKKLLKLKSEMEEKEYSLTKERDELMGKLRESEERSCELSCSVDLLKKRL 599
Db 541 MDVTEKLIBSKKLLKLKSEMEEKVYNLTREERDELIGLKSSEERSSELSCSVDLLKKRL 600
QY 600 DGBIEEVEREINRGRCKGSEFTCPEDNKIRELTLEIERLKKRLQQLVEVVEGDLMKTEYD 659
Db 601 DGBIEEVEREITRGRSRKGSSELTCPEDNKIKELTLEIERLKKRLQQLVEVVEGDLMKTEYD 660
QY 660 DQLEQKFRTEQDKANFLSQLEEBEIKHQMAKHAIEKGEAVSQEAELRHRFRLEEAKSRLD 719

Db	661	DQLEQKFRTEQDKANFLSQOLEEIKHQIAKNKAIEGCVVSQEAELRHRFRLEBAKSRDL	720
Qy	720	QAEVQALKEKIHLMNKKEDQLSQLQVDYSVLQORFMBEETKNKNMGREVNLTKLELSK	779
Db	721	KAEVQALKEKIHLMNKKEDQLSQLQVDYSVLQORFMBEENKNKNMGQEVNLTKLELSK	780
Qy	780	RYSRALRPSGNGRRMVDVPVASTGVQTEAVCGDAABEETPAVFIRKSFQEEHNHMSNLRQ	839
Db	781	RYSRALRPSVNGRRMVDVPVTSTGVQTDVAVSGEAABEETPAVFIRKSFQEEHNHMSNLRQ	840
Qy	840	VGLKKPMERSSVLDRYPPAANELTMRKSWIPWMRKRENGPSTPQEKGP RPNOGAGHPGEL	899
Db	841	VGLKKPVERSSVLDRYPPAANELTMRKSWIPWMRKRENGPSITQEKGPRTNSSPGHPGEV	900
Qy	900	VLAPKQGQPLHIRVTPDHENSTATLEITSPTSEEFSSSTTVIPTLGNQKPRITIIPSPNV	959
Db	901	VLSPKQGQPLHIRVTPDHENSTATLEITSPTSEEFSSSTTVIPTLGNQKPRITIIPSPNV	960
Qy	960	MSQPKXSADPTLGP RAMSPVTITTSIREKSPGEGRSFAFADRPASP IQIMTVSTSAAPTE	1019
Db	961	MPQKQKSGDTTLGP RAMSPVTITTSREKTPESGRGAFADRP TSPIQIMTVSTSAAPAE	1020
Qy	1020	IAVSPESQEVPMGRITILKVTPEKQTPAPVRKYNNSNANIITTEDNKIHIHLGSQFKRSPG	1079
Db	1021	IAVSPESQEMPGRITILKVTPEKQTPVTPVRKYNNSNANIITTEDNKIHIHLGSQFKRSPG	1080
Qy	1080	PAEGVSPVITVRPVNVTAEKEVSTGTVLRSPRNHLSRRPGASKVTSTITITPVTTSSTR	1139
Db	1081	TSGEGVSPVITVRPVNVTAEKEVSTGTVLRSPRNHLSRRPGASKVTSTITITPVTTSAR	1140
Qy	1140	GTQSVSGQDGSQRPTTRIPMSGKMGKAGPVVAASGAGNLTKFQPRAEQTOSMKIELKKS	1199
Db	1141	GTQSVSGQDGSQRPTTRIPMSGKMGKAGPVVAAPGAGNLTKFEPRAEQTOSMKIELKKS	1200
Qy	1200	AASSTASLGGGKG 1212	
Db	1201	AASSTTSLGGGKG 1213	
RESULT 3			
ID	Q86TC3	PRELIMINARY;	PRT; 1177 AA.
AC	Q86TC3;		
DT	01-JUN-2003 (TrEMBLrel. 24, Created)		
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Hypothetical protein DKFZp451B134.		
GN	DKFZp451B134.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,		
RA	Osanger A., Wiemann S.;		
RL	Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AL832009; CAD89912.1; -.		
DR	InterPro; IPR000533; Tropomyosin.		
DR	PRINTS; PR00194; TROPOMYOSIN.		
KW	Hypothetical protein.		
SQ	SEQUENCE 1177 AA; 134826 MW; D791D2A35FD10E4B CRC64;		
Query Match 89.7%; Score 5456.5; DB 4; Length 1177;			
Best Local Similarity 93.3%; Pred. No. 3.7e-207;			
Matches 1086; Conservative 36; Mismatches 41; Indels 1; Gaps 1;			
Qy	1	MRSRNQGGESSNGHVSCKPSSIISSDGGKGPSEDA-KKNKANRKEEDVMASGTIKRHLK	59
Db	1	MRSRNQGGESASDGHISCPKPSIIIGNAGEKSLSEDAKKRKNRKEEDVMASGTIVKRHLK	60
Qy	60	PSGESEKTKKSVELSKEDLIQLLSIMEGELQAREDIVIHLRTEKTKPEVLEAHYGSAP	119

Db	61	TSGECERKTKKSLSKEDLIQLLSIMEGELQAREDIVIHLKTEKTKPEVLEAHYGSAP	120
Qy	120	EKVLRLHRDAILAEKSIGEDVYEKPISELDRLEEKQKETYRRMLEQLLLAEKCHRTV	179
Db	121	EKVLRLHRDAILAEKSIGEDVYEKPISELDRLEEKQKETYRRMLEQLLLAEKCHRTV	180
Qy	180	YELENEKHKHTDYNMNSDDFTNLLEQERERLKKLLEQEKAYQARKEKENAKRLNKLDEL	239
Db	181	YELENEKHKHTDYNMNSDDFTNLLEQERERLKKLLEQEKAYQARKEKENAKRLNKLDEL	240
Qy	240	VKLKSFALMLVDERQMHIEQLGLSQSKVQDLTKLREEEEEKLKAVTYKSKEDRQKLLKLE	299
Db	241	VKLKSFALMLVDERQMHIEQLGLSQSKVQDLTKLREEEEEKLKAITSKSKEDRQKLLKLE	300
Qy	300	VDFEHKASRFSQHEEMNAKLANQESHNRQLRLKLVGLSQRIEELEETNKSLOKABEELQ	359
Db	301	VDFEHKASRFSQHEEMNAKLANQESHNRQLRLKLVGLTQRIEELEEINKLOKABEELQ	360
Qy	360	ELREKIAKGECSGSSIMAEVESLRKRVLEMEGKDEEITKTEAQCRELKKKLOEEHHKSKE	419
Db	361	ELRDKIAKGECSGSSIMAEVENLRKRVLEMEGKDEEITKTESQCRELKKKLOEEHHKSKE	420
Qy	420	LRLEVEKLOKRMSELEKLEEAFFSRSKSECTQLHLNLEKEKNLTKDLLNELEVYKSRVKEL	479
Db	421	LRLEVEKLOKRMSELEKLEEAFFSKSKSECTQLHLNLEKEKNLTKDLLNELEVYKSRVKEL	480
Qy	480	ECSESRLKEAELSKDDLTCLKSFTVMLVDERKNMMEKIKQEERKVDGLNKNFKVQGGKV	539
Db	481	ECSESRLKEAELSKDDLTCLKSFTVMLVDERKNMMEKIKQEERKVDGLNKNFKVQGGKV	540
Qy	540	MDVTEKLIIESKKLLKLKSEMEEKEYSLTKERDELGMGLRSEERSCELSVDLLKKRL	599
Db	541	MDVTEKLIIESKKLLKLKSEMEEKVYNLTREDELIGLKSEEEKSSELSVDLLKKRL	600
Qy	600	DGIEEVEREINRGSRCKSGSEFTCPEDNKIRELTLEIERLKKRLQQLLEVVEGDLMKTEDEY	659
Db	601	DGIEEVEREITRGRSRKSGSELTCPEDNKIKELTLEIERLKKRLQQLLEVVEGDLMKTEDEY	660
Qy	660	DQLEQKFRTEQDKANFLSQOLEEIKHQMAKHAIEKGEAVSQEAELRHRFRLEBAKSRDL	719
Db	661	DQLEQKFRTEQDKANFLSQOLEEIKHQIAKNKAIEGCVVSQEAELRHRFRLEBAKSRDL	720
Qy	720	QAEVQALKEKIHLMNKKEDQLSQLQVDYSVLQORFMBEETKNKNMGREVNLTKLELSK	779
Db	721	KAEVQALKEKIHLMNKKEDQLSQLQVDYSVLQORFMBEENKNKNMGQEVNLTKLELSK	780
Qy	780	RYSRALRPSGNGRRMVDVPVASTGVQTEAVCGDAABEETPAVFIRKSFQEEHNHMSNLRQ	839
Db	781	RYSRALRPSVNGRRMVDVPVTSTGVQTDVAVSGEAABEETPAVFIRKSFQEEHNHMSNLRQ	840
Qy	840	VGLKKPMERSSVLDRYPPAANELTMRKSWIPWMRKRENGPSTPQEKGP RPNOGAGHPGEL	899
Db	841	VGLKKPVERSSVLDRYPPAANELTMRKSWIPWMRKRENGPSITQEKGPRTNSSPGHPGEV	900
Qy	900	VLAPKQGQPLHIRVTPDHENSTATLEITSPTSEEFSSSTTVIPTLGNQKPRITIIPSPNV	959
Db	901	VLSPKQGQPLHIRVTPDHENSTATLEITSPTSEEFSSSTTVIPTLGNQKPRITIIPSPNV	960
Qy	960	MSQPKXSADPTLGP RAMSPVTITTSIREKSPGEGRSFAFADRPASP IQIMTVSTSAAPTE	1019
Db	961	MPQKQKSGDTTLGP RAMSPVTITTSREKTPESGRGAFADRP TSPIQIMTVSTSAAPAE	1020
Qy	1020	IAVSPESQEVPMGRITILKVTPEKQTPAPVRKYNNSNANIITTEDNKIHIHLGSQFKRSPG	1079
Db	1021	IAVSPESQEMPGRITILKVTPEKQTPVTPVRKYNNSNANIITTEDNKIHIHLGSQFKRSPG	1080
Qy	1080	PAAEVSPVITVRPVNVTAEKEVSTGTVLRSPRNHLSRRPGASKVTSTITITPVTTSSTR	1139
Db	1081	TSGEGVSPVITVRPVNVTAEKGVSTGTVLRSPRNHLSRRPGASKVTSTITITPVTTSAR	1140
Qy	1140	GTQSVSGQDGSQRPTTRIPMSK 1163	

Db 1141 GTQSVSGDSSQSRPTPTRIPMSK 1164

RESULT 4

Q9ULE5 PRELIMINARY; PRT; 1140 AA.

AC Q9ULE5;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein KIAA1275 (Fragment).

GN KIAA1275.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=20039619; PubMed=10574462;

RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro."

RL DNA Res. 6:337-345(1999).

DR EMBL; AB033101; BAA86589.1; -.

DR Genew; HGNC:21015; FILIP1.

DR InterPro; IPR000533; Tropomyosin.

DR PRINTS; PR00194; TROPOMYOSIN.

KW Hypothetical protein.

FT NON_TER

SQ SEQUENCE 1140 AA; 130283 MW; 79EC2AB496556415 CRC64;

Query Match 87.9%; Score 5349; DB 4; Length 1140;

Best Local Similarity 93.8%; Pred. No. 6.2e-203;

Matches 1064; Conservative 32; Mismatches 38; Indels 0; Gaps 0;

Qy 79 LIQLLSIMBEGELQAREDVIHMLRTEKTPKPEVLEAHYGSABPEKVLRLVLRDAILAQEKSI 138

Db 7 IVSRLLHFLHLLQAREDVIHMLKTEKTPKPEVLEAHYGSABPEKVLRLVLRDAILAQEKSI 66

Qy 139 GEDVYEKPISELDRLEEKQETRYRMLEQLLLAEKCHRTTVYELENEKHKHTDYMNKSD 198

Db 67 GEDVYEKPISELDRLEEKQETRYRMLEQLLLAEKCHRTTVYELENEKHKHTDYMNKSD 126

Qy 199 FTNLLQEERERLKLLEQEKAYQARKEKENAKRLNKLRLDELVLKLSFALMLVDERQMHIE 258

Db 127 FTNLLQEERERLKLLEQEKAYQARKEKENAKRLNKLRLDELVLKLSFALMLVDERQMHIE 186

Qy 259 QLGLSQKQVDLTQKLREEEKLKAVTYKSKEDRQKLKLEVDVFEHKASRFSQEEHEMNA 318

Db 187 QLGLSQKQVDLTQKLREEEKLKAITSKSKEDRQKLKLEVDVFEHKASRFSQEEHEMNA 246

Qy 319 KLANQESHNRQLRLKLVLSQRIEELETNKSLOKAEELQELREKIAKGECGNSSLMAB 378

Db 247 KLANQESHNRQLRLKLVLTQRIEELETNKNLQKAEELQELRDKIAKGECGNSSLMAB 306

Qy 379 VESLRKRVLEMEGKDEETKTTEAQCRELKKLQEEHHSKELRLEVEKLQKRMSELEKLE 438

Db 307 VENLRKRVLEMEGKDEETKTESQCRELRKKLQEEHHSKELRLEVEKLQKRMSELEKLE 366

Qy 439 EAFSRKSSECTQLHLNLEKEKNLTKDLNLELEVVKSRVKELECSERLEKAELSLKDDLT 498

Db 367 EAFSKSSECTQLHLNLEKEKNLTKDLNLELEVVKSRVKELECSERLEKAELSLKDDLT 426

Qy 499 KLKSFTVMLVDERKNMMEKIKQEEERKVDGLNKNFKVEQGVMDVTEKLIIESKLLKLKS 558

Db 427 KLKSFTVMLVDERKNMMEKIKQEEERKVDGLNKNFKVEQGVMDVTEKLIIESKLLKLKS 486

Qy 559 EMBEKEYSLTKERDELMGKLRSEERSCELSVLLKKRLDGIIEVEREINRGRSKGS 618

Db 487 EMEEKVYNLTRERDELIGLKSSEEEKSELSCSVDLLKKRLDGIIEVEREITRGRSKGS 546

Qy 619 EFTCPEDNKKIRELTLEIERLKKRLQQLLEVVEGDLMKTEDEYDQLEQKFRTEQDKANFLSQ 678

Db 547 ELTCEPDNKKIKELTLEIERLKKRLQQLLEVVEGDLMKTEDEYDQLEQKFRTEQDKANFLSQ 606

Qy 679 QLEEIKHQMAKHKAIEKGEAVSQEAELRHRFRLEEAKSRDLQAEVQALKKEIHELMNKED 738

Db 607 QLEEIKHQIAKNKAIEKGEVVSQEAELRHRFRLEEAKSRDLQAEVQALKKEIHELMNKED 666

Qy 739 QLSQLQVDYSVLQORFMEETKKNMGRVNLTKLELSKRYSRALRPSGNRRMVDVP 798

Db 667 QLSQLQVDYSVLQORFMEENKKNMGQEVNLTKLELSKRYSRALRPSVNGRRMVDVP 726

Qy 799 VASTGVQTEAVCGDAEEETPAVFIRKSFQBEENHIMSILRQVGLKKPMERSSVLDYPPA 858

Db 727 VTSTGVQTDVAVSGEAAEEETPAVFIRKSFQBEENHIMSILRQVGLKKPVERSSVLDYPPA 786

Qy 859 ANELTMRKSWIPWMRKRENGPSTPOEKGPRPNQAGHPGELVLAPKQGOPLHIRVTPDHE 918

Db 787 ANELTMRKSWIPWMRKRENGPSTPOEKGPRPNQAGHPGELVLAPKQGOPLHIRVTPDHE 846

Qy 919 NSTATLEITSPTESEFFSSTTVIPTLGNQKPRITIIIPSPNVMSQPKSADPTLGPERRAMS 978

Db 847 NSTATLEITSPTESEFFSSTTVIPTLGNQKPRITIIIPSPNVMPQKQSGDTTLGPERRAMS 906

Qy 979 PVTITITISREKSPGGRSAFADRPASPIQIMTVSTSAAPTEIAVSPESQEVPMGRITLKV 1038

Db 907 PVTITITISREKTPESGRGAFADRPTSPIQIMTVSTSAAPTEIAVSPESQEMPMPGRITLKV 966

Qy 1039 TPEKQTVPAVPVRKYNSNANIITTEDNKIHLGSGQFKRSPGPAAGVSPVITVRPVNVTA 1098

Db 967 TPEKQTVPTVPRKYNSNANIITTEDNKIHLGSGQFKRSPGTSGEVSPVITVRPVNVTA 1026

Qy 1099 EKEVSTGTVLRSPRNHLSRPGASKVTSTITITPVTTSSTRGTQSVSGDQSSQRPPTPR 1158

Db 1027 EKEVSTGTVLRSPRNHLSRPGASKVTSTITITPVTTSSTRGTQSVSGDQSSQRPPTPR 1086

Qy 1159 IPMSKGMKAGKPVVAASGAGNLTQFQRAETQSMKIELKKSAASTASLGGGKG 1212

Db 1087 IPMSKGMKAGKPVVAAPGAGNLTQFQRAETQSMKIELKKSAASTSLGGGKG 1140

RESULT 5

Q8JZS5 PRELIMINARY; PRT; 965 AA.

AC Q8JZS5;

DT 01-OCT-2002 (TrEMBLrel. 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Filamin-interacting protein S-FILIP.

GN S-FILIP.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Wistar; TISSUE=Forebrain;

RX MEDLINE=22100428; PubMed=12055638;

RA Nagano T., Yoneda T., Hatanaka Y., Kubota C., Murakami F., Sato M.;

RT "Filamin A-interacting protein (FILIP) regulates cortical cell migration out of the ventricular zone.";

RL Nat. Cell Biol. 4:495-501(2002).

DR EMBL; D87257; BAC00852.1; -.

DR InterPro; IPR000533; Tropomyosin.

DR PRINTS; PR00194; TROPOMYOSIN.

SQ SEQUENCE 965 AA; 109188 MW; 50352DF4F8236B40 CRC64;

Query Match 79.5%; Score 4838; DB 11; Length 965;

Best Local Similarity 100.0%; Pred. No. 6.8e-183;

Matches 965; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 248 MLVDERQMHIEQLGLSQKQVDLTQKLREEEKLKAVTYKSKEDRQKLKLEVDVFEHKAS 307

Db 1 MLVDERQMHIQLGLSQKVQDLTQKLREEEEKLKAVTYKSKEDRQKLLKLEVDPEHKAS 60

QY 308 RFSQHEEMNAKLANQESHNRQLRLKLVGLSORIEELEETNKSLOKAEELQELREKIAK 367

Db 61 RFSQHEEMNAKLANQESHNRQLRLKLVGLSORIEELEETNKSLOKAEELQELREKIAK 120

QY 368 GECGNSSLMAEVESLRKRVLEMEGKDEEITKTEAQCRELKKLQEEEEHHSKELRLEVEKL 427

Db 121 GECGNSSLMAEVESLRKRVLEMEGKDEEITKTEAQCRELKKLQEEEEHHSKELRLEVEKL 180

QY 428 QKRMSELEKLEEAFFSKSECTQLHLNLEKEKNLT KDLLNELEVVKSRVKELECESRLE 487

Db 181 QKRMSELEKLEEAFFSKSECTQLHLNLEKEKNLT KDLLNELEVVKSRVKELECESRLE 240

QY 488 KAE LSLKDDLT KLKSF T VMLVDERKNMMEKIQEERKVDGLNKNFKVQGGKVMVDVTEKLI 547

Db 241 KAE LSLKDDLT KLKSF T VMLVDERKNMMEKIQEERKVDGLNKNFKVQGGKVMVDVTEKLI 300

QY 548 EESKLLK LKSEMEEKEYSLTKERDELMGKLRSERSESCSVDLLKKRLDGIIEEVER 607

Db 301 EESKLLK LKSEMEEKEYSLTKERDELMGKLRSERSESCSVDLLKKRLDGIIEEVER 360

QY 608 EINRGRSCKGSEFTCPEDNKIRELTLEIERLKKRLQQLLEVVEGDLMKTEYDQLEQKFR 667

Db 361 EINRGRSCKGSEFTCPEDNKIRELTLEIERLKKRLQQLLEVVEGDLMKTEYDQLEQKFR 420

QY 668 TEQDKANFLSQOLEEIKHQMAKHAIEKGEAVSQAE LRRFRLEEA KSRDLQAEVQALK 727

Db 421 TEQDKANFLSQOLEEIKHQMAKHAIEKGEAVSQAE LRRFRLEEA KSRDLQAEVQALK 480

QY 728 EKIHELMNKEDQLSQLDVYSVLQQRFMEEETKNKMGMREVNLTKELSKRYSRALRP 787

Db 481 EKIHELMNKEDQLSQLDVYSVLQQRFMEEETKNKMGMREVNLTKELSKRYSRALRP 540

QY 788 SGNRRMVDVPVASTGVQTEAVCGDAAEEETPAVFIKRSFQEEHIMS NLRQVGLKKPME 847

Db 541 SGNRRMVDVPVASTGVQTEAVCGDAAEEETPAVFIKRSFQEEHIMS NLRQVGLKKPME 600

QY 848 RSSVLDRYP PPAANELTMRKSWIPWMRKRENGPSTPQEKGRPNQAGHPGELVLAPKQGQ 907

Db 601 RSSVLDRYP PPAANELTMRKSWIPWMRKRENGPSTPQEKGRPNQAGHPGELVLAPKQGQ 660

QY 908 PLHIRVTPDHENSTATLEITSPTSEEFFSSTTVIPTLGNQKPRITIIIPSPNVMSQKPKSA 967

Db 661 PLHIRVTPDHENSTATLEITSPTSEEFFSSTTVIPTLGNQKPRITIIIPSPNVMSQKPKSA 720

QY 968 DPTLGP ERAMSPVTITISREKSP EGGRSAFADRPASPIQIMTVSTSAAPTETIAVSPESQ 1027

Db 721 DPTLGP ERAMSPVTITISREKSP EGGRSAFADRPASPIQIMTVSTSAAPTETIAVSPESQ 780

QY 1028 EVPMGRTILKVTP EKQTV PAVRKYNSNANIITTEDNKIHIHLG SQFKRSPGPAAGVSP 1087

Db 781 EVPMGRTILKVTP EKQTV PAVRKYNSNANIITTEDNKIHIHLG SQFKRSPGPAAGVSP 840

QY 1088 VITVRPNVNTAEKEVSTGTVLRSRPNHLSSRPGASKVTSTITITPVTTS TRGTQSVSGQ 1147

Db 841 VITVRPNVNTAEKEVSTGTVLRSRPNHLSSRPGASKVTSTITITPVTTS TRGTQSVSGQ 900

QY 1148 DGSSQRPTPTRI PMSKGMKAGKPVVAASGAGNLTKFPQRAETQSMKIELKSAASSTASL 1207

Db 901 DGSSQRPTPTRI PMSKGMKAGKPVVAASGAGNLTKFPQRAETQSMKIELKSAASSTASL 960

QY 1208 GGGKG 1212

Db 961 GGGKG 965

RESULT 6
Q96SK6

ID Q96SK6 PRELIMINARY; PRT; 965 AA.

AC Q96SK6;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Hypothetical protein FLJ14799.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA ISOGAI T., Ota T., Hayaashi K., Sugiyama T., Otsuki T., Suzuki Y.,

RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,

RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,

RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,

RA Masuho Y., Kanehori K.;

RT "NEDO human cDNA sequencing project.";

RL Submitted (MAY-2001) to the EMBL/GenBank/DBDJ databases.

DR EMBL; AK027705; BAB55310.1; -.

DR InterPro; IPR000533; Tropomyosin.

DR PRINTS; PR00194; TROPOMYOSIN.

KW Hypothetical protein.

SQ SEQUENCE 965 AA; 109200 MW; 25EB750B47E42433 CRC64;

Query Match 74.7%; Score 4541; DB 4; Length 965;

Best Local Similarity 93.7%; Pred. No.3.3e-171;

Matches 904; Conservative 28; Mismatches 33; Indels 0; Gaps 0;

QY 248 MLVDERQMHIQLGLSQKVQDLTQKLREEEEKLKAVTYKSKEDRQKLLKLEVDPEHKAS 307

Db 1 MLVDERQMHIQLGLSQKVQDLTQKLREEEEKLKAITSKSKEDRQKLLKLEVDPEHKAS 60

QY 308 RFSQHEEMNAKLANQESHNRQLRLKLVGLSORIEELEETNKSLOKAEELQELREKIAK 367

Db 61 RFSQHEEMNAKLANQESHNRQLRLKLVGLSORIEELEETNKNLQKAEELQELRDKIAK 120

QY 368 GECGNSSLMAEVESLRKRVLEMEGKDEEITKTEAQCRELKKLQEEEEHHSKELRLEVEKL 427

Db 121 GECGNSSLMAEVENLRKRVLEMEGKDEEITKTESQCRELKKLQEEEEHHSKELRLEVEKL 180

QY 428 QKRMSELEKLEEAFFSKSECTQLHLNLEKEKNLT KDLLNELEVVKSRVKELECESRLE 487

Db 181 QKRMSELEKLEEAFFSKSECTQLHLNLEKEKNLT KDLLNELEVVKSRVKELECESRLE 240

QY 488 KAE LSLKDDLT KLKSF T VMLVDERKNMMEKIQEERKVDGLNKNFKVQGGKVMVDVTEKLI 547

Db 241 KAE LSLKDDLT KLKSF T VMLVDERKNMMEKIQEERKVDGLNKNFKVQGGKVMVDVTEKLI 300

QY 548 EESKLLK LKSEMEEKEYSLTKERDELMGKLRSERSESCSVDLLKKRLDGIIEEVER 607

Db 301 EESKLLK LKSEMEEKVYNLTRERDELIGLKSEEEKSELSCSVDLLKKRLDGIIEEVER 360

QY 608 EINRGRSCKGSEFTCPEDNKIRELTLEIERLKKRLQQLLEVVEGDLMKTEYDQLEQKFR 667

Db 361 EITRGRSRKGSBLT CPEDNKIKELTLEIERLKKRLQQLLEVVEGDLMKTEYDQLEQKFR 420

QY 668 TEQDKANFLSQOLEEIKHQMAKHAIEKGEAVSQAE LRRFRLEEA KSRDLQAEVQALK 727

Db 421 TEQDKANFLSQOLEEIKHQIAKNAIEKGEVVSQAE LRRFRLEEA KSRDLKAGVQALK 480

QY 728 EKIHELMNKEDQLSQLDVYSVLQQRFMEEETKNKMGMREVNLTKELSKRYSRALRP 787

Db 481 EKIHELMNKEDQLSQLDVYSVLQQRFMEEENKNKMGMQEVNLTKELSKRYSRALRP 540

QY 788 SGNRRMVDVPVASTGVQTEAVCGDAAEEETPAVFIKRSFQEEHIMS NLRQVGLKKPME 847

Db 541 SVNGKRMVDVPVTSTGVQTDAVSGEAAEEETPAVFIKRSFQEEHIMS NLRQVGLKKPVE 600

QY 848 RSSVLDRYP PPAANELTMRKSWIPWMRKRENGPSTPQEKGRPNQAGHPGELVLAPKQGQ 907

Db 601 RSSVLDRYP PPAANELTMRKSWIPWMRKRENGPSITQEKGRPTNSSPGHPGGEVVLSPKQGQ 660

QY 908 PLHIRVTPDHENSTATLEITSPTSEEFFSSTTVIPTLGNQKPRITIIIPSPNVMSQKPKSA 967

Db 661 PLHIRVTPDHENSTATLEITSPTSEEFFSSTTVIPTLGNQKPRITIIIPSPNMPQKQKSG 720

Db 95 MRSRNOGGESSNGHVSCPKPSIISSDGGKGPSEDAKKNKANRKGEDDVMSGTVKRHLK 154
QY 60 PSGESEKTKKSVELSKEDLIQLLSIMEGELOAREDVIHMLRTEKTKPEVLEAHYGSAP 119
Db 155 PSGESEKTKKPLELSKEDLIQLLSIMEGELOAREDVIHMLKTEKTKPEVLEAHYGSAP 214
QY 120 EKVLRLVLRDAILAQAQKSIGEDVYEKPISELDRLEEKQKETYRRMLQQLLLAEKCHRTV 179
Db 215 EKVLRLVLRDAILAQAQKSIGEDVYEKPISELDRLEEKQKETYRRMLQQLLLAEKCHRTV 274
QY 180 YELENEKHKHTDYMNKSDDDFTNLLQEQRERLKKLLEQEKAYQARKEKENAKRLNKLDEL 239
Db 275 YELENEKHKHTDYMNKSDDDFTNLLQEQRERLKKLLEQEKAYQARKEKENAKRLNKLDEL 334
QY 240 VKLKSFALMLVDERQMHIEQLGLSQKVQDLTQKLREEBEKLVAVTYKSKEDRQKLLKLE 299
Db 335 VKLKSFALMLVDERQMHIEQLGLSQKVQDLTQKLREEBEKLVAVTYKSKEDRQKLLKLE 394
QY 300 VDFEHKASRFSQEHEEMNAKLANQESHNRQLRLKLVGLSQRIELEEETNKSLOKABEELQ 359
Db 395 VDFEHKASRFSQEHEEMNAKLANQESHNRQLRLKLVGLSQRIELEEETNKSLOKABEELQ 454
QY 360 ELREKIAKGECCGNSSLMAEVESLRKRVLEMEGKDEEITKTEAQCRELKKLQEEHHKSKE 419
Db 455 ELRDKIAKGECCGNSSLMAEVESLRKRVLEMEGKDEEITKTEAQCRELKKLQEEHHKSKE 514
QY 420 LRLEVEKLOKRMSELEKLEBAFSAFSSKSECTQLHLNLEKEKNLTKDLLNELEVVKSRVKEL 479
Db 515 LRLEVEKLOKRMSELEKLEBAFSAFSSKSECTQLHLNLEKEKNLTKDLLNELEVVKSRVKEL 574
QY 480 ECSESRLKAEKLSKDDLTKLKSFTVMLVDERKNMM 515
Db 575 ECSESRLKAEKLSKDDLTKLKSFTVMLVDERKNMM 610

RESULT 9
Q8N8B9
ID Q8N8B9 PRELIMINARY; PRT; 512 AA.
AC Q8N8B9;
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ39702 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Small intestine;
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK097021; BAC04928.1; -;
DR InterPro; IPR000533; TROPOMYOSIN.
DR PRINTS; PR00194; TROPOMYOSIN.
KW Hypothetical protein.
FT NON TER 512 512
SQ SEQUENCE 512 AA; 60035 MW; 41F3B8A2FBD28E20 CRC64;
Query Match 39.6%; Score 2406.5; DB 4; Length 512;
Best Local Similarity 94.1%; Pred. No. 1.8e-87;
Matches 482; Conservative 18; Mismatches 11; Indels 1; Gaps 1;
QY 1 MRSRNOGGESSNGHVSCPKPSIISSDGGKGPSEDA-KKNKANRKEEDVMSGTVKRHLK 59

Db 1 MRSRNOGGESASDGHISCPKPSIIGNAGEKSLSEDAKKKKSNRKEDDVMSGTVKRHLK 60
QY 60 PSGESEKTKKSVELSKEDLIQLLSIMEGELOAREDVIHMLRTEKTKPEVLEAHYGSAP 119
Db 61 TSCECERKTKKSLELSKEDLIQLLSIMEGELOAREDVIHMLKTEKTKPEVLEAHYGSAP 120
QY 120 EKVLRLVLRDAILAQAQKSIGEDVYEKPISELDRLEEKQKETYRRMLQQLLLAEKCHRTV 179
Db 121 EKVLRLVLRDAILAQAQKSIGEDVYEKPISELDRLEEKQKETYRRMLQQLLLAEKCHRTV 180
QY 180 YELENEKHKHTDYMNKSDDDFTNLLQEQRERLKKLLEQEKAYQARKEKENAKRLNKLDEL 239
Db 181 YELENEKHKHTDYMNKSDDDFTNLLQEQRERLKKLLEQEKAYQARKEKENAKRLNKLDEL 240
QY 240 VKLKSFALMLVDERQMHIEQLGLSQKVQDLTQKLREEBEKLVAVTYKSKEDRQKLLKLE 299
Db 241 VKLKSFALMLVDERQMHIEQLGLSQKVQDLTQKLREEBEKLVAVTYKSKEDRQKLLKLE 300
QY 300 VDFEHKASRFSQEHEEMNAKLANQESHNRQLRLKLVGLSQRIELEEETNKSLOKABEELQ 359
Db 301 VDFEHKASRFSQEHEEMNAKLANQESHNRQLRLKLVGLTQRIELEEETNKNLQKABEELQ 360
QY 360 ELREKIAKGECCGNSSLMAEVESLRKRVLEMEGKDEEITKTEAQCRELKKLQEEHHKSKE 419
Db 361 ELRDKIAKGECCGNSSLMAEVENLRKRVLEMEGKDEEITKTESQCRELKKLQEEHHKSKE 420
QY 420 LRLEVEKLOKRMSELEKLEBAFSAFSSKSECTQLHLNLEKEKNLTKDLLNELEVVKSRVKEL 479
Db 421 FRLEVEKLOKRMSELEKLEBAFSAFSSKSECTQLHLNLEKEKNLTKDLLNELEVVKSRVKEL 480
QY 480 ECSESRLKAEKLSKDDLTKLKSFTVMLVDER 511
Db 481 ECSESRLKAEKLSKDDLTKLKSFTVMLVDER 512

RESULT 10
Q8N6Z0
ID Q8N6Z0 PRELIMINARY; PRT; 893 AA.
AC Q8N6Z0;
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TremBLrel. 22, Last annotation update)
DE Similar to downregulated in ovarian cancer 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027860; AAH27860.1; -;
SQ SEQUENCE 893 AA; 102232 MW; 561BE17D6C10FDC4 CRC64;
Query Match 31.7%; Score 1926; DB 4; Length 893;
Best Local Similarity 46.0%; Pred. No. 2.7e-68;
Matches 424; Conservative 168; Mismatches 266; Indels 64; Gaps 14;
QY 248 MLVDERQMHIEQLGLSQKVQDLTQKLREEBEKLVAVTYKSKEDRQKLLKLEVDFEHKAS 307
Db 1 MVDDEQORLTAQLTLQRQKIQLTNNAKETHYKLALAEARVQEEEQKATRLKELQQTQT 60
QY 308 RFSQEHEEMNAKLANQESHNRQLRLKLVGLSQRIELEEETNKSLOKABEELQELREKIAK 367
Db 61 KFHQDQDTIMAKLTNEDSQNRQLQKLAALSQIDLEETNRSRLRKAEEELQDIKEKISK 120
QY 368 GECGNSSLMAEVESLRKRVLEMEGKDEEITKTEAQCRELKKLQEEHHKSKELRLEVEKL 427
Db 121 GEYGNAGIMAEVEELRKRVLDMEGKDEELIKMEEQCRDLNKRLERETQLQSKDFLEVEKL 180
QY 428 QKRMSELEKLEBAFSAFSSKSECTQLHLNLEKEKNLTKDLLNELEVVKSRVKELECSERLE 487

Db	181	SKRIMALEKLEDAFNKSKQECYSLKCNLEKERMTTKQLSQELES�KVRKELEAIESRLE	240
QY	488	KAEĹSLKDDLTKĹKSFTVMLVDERKNMMEKIĲOEERKVDGLNKNFKVEQKVMDEVTEKLI	547
Db	241	KTEFTĹKEDĹTKĹKTLTVMFVDERKTMSEKĹKTKEDĹĲAASSQĹVEQKVTTVTEKLI	300
QY	548	EESKĹĹKĹKSEMEEEKYSLTKERDEĹMGĹRSEBERSCELSCSVĹĹKĹRĹDĹGIEEVER	607
Db	301	EETKĹALKSKTDVEEKMYSVTKERDDĹKĹKĹKAEEEKGNĹĹSRVNMĹKĹRĹQSLEAIEK	360
QY	608	EINRGR-----SCKGSEFTCPEDNKIREĹTĹIEIRĹKĹRĹQĹLEVVEGĹMLKTEDEYDQĹE	663
Db	361	DFĹKĹKĹNQDSGKSTTĹĹHQENNĹKĹELSQEVEĹRĹKĹĹKĹDMĲAIEDĹMLKTEDEYETĹE	420
QY	664	QĲFRTEQĲKANFLSQĹLEEIKHQMAKHĲAIEĲGEAVSQEĹELRHRFRLEEAKSRĹQAEV	723
Db	421	RRYANERĲAQĹFLSKEĹEHVKMEĹAKYĹĹAEĲTE-TSHEQWĹFKRĹQEEEAĲSGĹLSREV	479
QY	724	QĹĲKEĲIHEĹMNĲEDQĹSQĹVDYSVĹQRPMEEEETKĹKNMGREVLNĹTKĹEĹLSKRYSR	783
Db	480	DĹĲKEĲIHEYMĲATEDĹIĲHQGDHSVĹQĹKĹNĹQOENRĹRĹDĹGREIENĹTKĹELERYRHF	539
QY	784	ĹĹRPSGNGRMVĹVPVASTGVQTEAVCGDĹAABEETP-----AVFIRKSFOE-ENH	832
Db	540	SLRPSĹNGRISDPQVFSKEVQTEAV-----DNEPPDYĲSLIPLERAVINGQĹYEESENQ	594
QY	833	IMSĹNĹRQVĹĹĲĲPMERSSVĹ-----DRYPĲAĹNELTMRĲSWIĲWMRĲRE-----NGPSTPQ	883
Db	595	D-----EDPNDEGSVĹSFĲCSQSTĲCPVĹ-----RĲĹWĹPMĲKSKEGHĹQNG-----	636
QY	884	EĲGPRNQAGĲPGĹELVĹĹPKQOĲĹHIRVTPDHENSTATĹEITSPĲSE--EFFSSTTVI	941
Db	637	ĲMQTKĲNANFVQĲGĹVĹSHTPQĲĹHKVTPDHVQNTATĹEITSPĲTESPHSYTSTAVI	696
QY	942	PTĹGNQĲPRITĹIĲSPNVMSQĲPK-SADPTĹĲGERAMSPVTĹTĹISREĲSPĲGRSĹAFAD	1000
Db	697	PNCGTPĲQRTĹIĹQNASITĲPVĲSKSTSTEDĹMNĹEQGMSPIĲWĲTĲFARAQĲTPĲSCĲSLTPE	756
QY	1001	RPASPIQIMTVSTSAAPTEĹIAVSPĲESQĲVPMGRTĹĹKVTPEĲQTVĲAPVRĲKYNMĹNĹIIT	1060
Db	757	RTWSPIQVĹAVTGSASSPEQGRSPĲEPTĹEISAKĹĹAIFRVSPDRQSSWQFQRSNNSSSVIT	816
QY	1061	TEDNĲIHĹĹGQFKRSPGĲAĲGVSPVITVRPVNVĲTAEĲEVSTGTVĹRSRPNĹHĹSSRPĲ	1120
Db	817	TEDNĲIHĹĹGSPYMQA--VASPVRPASPSĹPLQDNRTQĹĹNGĹLNKT-----	863
QY	1121	ĲSKVTSTĹITĹPTVTSSTRGTQ 1142	
Db	864	TĹNĲVTSSITĹTĲATPLPRQSQ 885	
RESULT 11			
Q8IUM3			
ID	Q8IUM3	PRELIMINARY;	PRT; 764 AA.
AC	Q8IUM3;		
DT	01-MAR-2003 (TReMBLrel. 23, Created)		
DT	01-MAR-2003 (TReMBLrel. 23, Last sequence update)		
DT	01-JUN-2003 (TReMBLrel. 24, Last annotation update)		
DE	GPBP-interacting protein 90.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Skeletal muscle;		
RA	Revert-Ros F., Raya A., Granero F., Saus J.;		
RT	"Cloning and expression of GIP90, a protein down-regulated in cancer		
RT	cells which interacts with GPBP.";		
RL	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF329032; AAN16206.1; --		
DR	InterPro; IPR000533; Tropomyosin.		

DR	PRINTS; PR00194; TROPOMYOSIN.		
SQ	SEQUENCE 764 AA; 89785 MW; DB41B4CB4EEFA085 CRC64;		
	Query Match 30.1%; Score 1828; DB 4; Length 764;		
	Best Local Similarity 50.7%; Pred. No. 1.7e-64;		
	Matches 386; Conservative 144; Mismatches 210; Indels 22; Gaps 7;		
QY	1	MRSRNOGGESSNGHVSCPSSIISSDGKGSPSEDĲKNKANRKEE-DVMASGTIKRĹĹK	59
Db	1	MRSR--GSDTEGSAQĲFRHT-----ĲGHSFQĲPNMĲHRQDĲDSPESDV---IL	48
QY	60	PSGESEK----KTKKSVELSKEDĹIQLĹSIMEGELQĹAREDVIHMLRTEĲTĲEVĹEAHYG	115
Db	49	PCĲKAĲPHSGNGHQAEDĹSRDDĹĹFLĹSILEGELQĹARDEVIGĹĲAEĲMDĹĹĹEAQYG	108
QY	116	SAEPEĲVLĹRĹDAĹĹAQĲKSIGEDVYĲĲPISELDRLEEQĲKĲTYRRMĹEQĹĹĹAEĲCH	175
Db	109	FVTPĲKVĹEĹALQĲDAFQĲKSTPQWQEDĹYĲĲPMNĹLDĲVVEĲĲKESYRRĹILQĹĹVAĲKS	168
QY	176	RRTVYELĲNEĲKHĲHTDYMNKSDDTNĹLEQĲERĹRĲĲĹEĲQĲKĹYQĲARKEĲENĲRĹNĹKĹ	235
Db	169	RQĹILEEEĲKRĲĲMEĲKSDEFĹCLĹEQĲCERĲĲĹĹIDQĹĲKSQEEĲKEĲKRVTTĹ	228
QY	236	RĹELVĲĲSFĹMLVDERQMHIEQĹĲQSQVQDĹTQĲREBEĲKĹKAVTYĲSKEDRQĲĹ	295
Db	229	KEĹTĲĲKSĲALMVVDEQĲRLTAQĹTĹRQĲĲELTĹTNAĲETHĲĲĹĹAEĲARVQEEĲQĲA	288
QY	296	ĹĲEVDĲFEĲKĲSRFSQĲHEEĲNĲANQĲESHNRQĲRLĲVĹĹSQRIEĹEĹEETNĲKSĹQĲAE	355
Db	289	TRĹEKĹEQĲTQTKFHQDQDTĹMAĲĹTNEĹDSNRQĲQĲĲĹĹSRQĹIDEĹEETNRĲSRĲAE	348
QY	356	EĹQĹELREĲĲAKGECGNSSĹMAĲVESĹRĲVĹMEĲGĲDEĹITĲTEĲAQĲRELĲĲĲQEEĲH	415
Db	349	BEĹQDIĲKEĲĲSGEYGNĲAGĹMAĲVEĹRĲRVĹDMĲEGĲDEĹĹĲMEĲQĲRĹNĲRĹERĹTĹ	408
QY	416	HSĲELRĹVEĲĲQĲRMSEĲĲEEĲAFSRSKSEĲTQĲĹĹNĹLEĲENĹTKĲĹĹNEĹEVVĲKS	475
Db	409	QSKDFĲĹEVEĲĲSKRĹMAĲĲLEĲDAFNĲSKQĲCYSLĲCNĲEKERMTTKQĹSQEĹESĲKVR	468
QY	476	VĲEĲCESĲSRĲEĲĲSLĲDDĹTKĲĲKSFTVMLVDERĲNMMEĲĲĲQĲERĲVDĲĲNĲFKVE	535
Db	469	ĲKEĹAĹIESRĲEĲTEFTĲĲEDĹTKĲĲTLTVMFVDERĲKTMSEĲĲKĲTKEDĲĲAASSQĲQVE	528
QY	536	QĲVMDEVTEĲĲEESĲĲĲĲKĲSEMEĲEĲEYSĹTKERDEĹMGĲRSEERSCELSCSVĹĹĹ	595
Db	529	QNKVTTVTEĲĲIEETĲRĲĲSKTDVEĲĲMYSVTKERDDĲĲNĲĲAEĲEĲGNĹĹĹSRVNMĹ	588
QY	596	ĲĲĹDĲGIEĲEVERĲINRGR---SĲĲGSEFTCPEDNKIREĹTĹIEIRĲĲKĲRĲQĹLEVVEGD	651
Db	589	ĲNĹRĲSQSLEĹAĲEĲDĲĲNĲNQNDSĲKSTTĹĹHQENNĲĲĲELSQEVEĹRĲĲĲĲĲĲĲĲĲĲĲĲ	648
QY	652	ĹMĲTEDEYDQĲEQFRTEQĲKANFLSQĹLEEIKHQMAKHĲAIEĲGEAVSQEĹELRHRFRĹ	711
Db	649	ĹMĲTEDEYETĹRĹRYANERDĲAQĹLSKEĹEHVKMEĹAKYĲĲAEĲTE-TSHEQWĹFKRĲQE	707
QY	712	EEĲKSRĹDĲQAEVQĲĲKEĲIHEĹMNĲEDQĹSQĹQVDYSVĹQQR 753	
Db	708	EEĲKSGĹLSREVĹDĲĲKEĲIHEYMĲATEDĹIĲHQGDHSVĹQĲK 749	
RESULT 12			
Q13597			
ID	Q13597	PRELIMINARY;	PRT; 752 AA.
AC	Q13597;		
DT	01-NOV-1996 (TReMBLrel. 01, Created)		
DT	01-NOV-1996 (TReMBLrel. 01, Last sequence update)		
DT	01-JUN-2003 (TReMBLrel. 24, Last annotation update)		
DE	DOC1.		
GN	DOC1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		

```

[1]
RN  SEQUENCE FROM N.A.
RP  MEDLINE=94148289; PubMed=8314147;
RX  Mok S.C., Wong K.K., Chan R.K., Lau C.C., Tsao S.W., Knapp R.C.,
RA  Berkowitz R.S.;
RT  "Molecular-cloning of differentially expressed genes in human
RL  epithelial ovarian cancer.";
RN  Gynecol. Oncol. 52:247-252(1994).
[2]
RN  SEQUENCE FROM N.A.
RP  Wong K.K., Mok S.C.;
RT  "Cloning and sequencing of full length Doc1 and Doc2 mRNAs.";
RL  Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR  EMBL; U53445; AAA98972.1; -.
DR  GO; GO:0016459; C:myosin; NAS.
SQ  SEQUENCE 752 AA; 87141 MW; D763B4E2806A3B0B CRC64;

Query Match      27.1%; Score 1645.5; DB 4; Length 752;
Best Local Similarity 46.3%; Pred. No. 2.5e-57;
Matches 363; Conservative 140; Mismatches 216; Indels 65; Gaps 13;

QY  248 MLVDERQMHIEQLGLOSQKVDLTQKLREEEEKLKAVTYKSKEDROKLLKLVDFEHKAS 307
Db  1 MVVDEQQRLLTAQLTQRQIKELTTNAKETHHTKLAABEARVQEEEQKATRLKELQTQT 60
QY  308 RFSQHEEMNAKLANQESHNRQLRLKLVGLSQRIEELEETNKSLOKAEELQELREKIAK 367
Db  61 KFHQDQDTIMAKLTNDSQNRQLQOKLAALSQRIDLEETNRSRLKAAEELQDIKEKISK 120
QY  368 GECGNSSLMAEVESLRKRVLEMEGKDEEITKTEAQCRELKKKLQEEHHHSKEIRLEVEKL 427
Db  121 GEYGNAGIMAEV-----EELIKMEEQCRDLNKRLERETLOSKDPKLEVEKL 166
QY  428 QKRMSELEKLEEAFFSKSECTQLHLNLEKEKNLTLDLLNELEVVKSRVKELECSERLE 487
Db  167 SKRIMALEKLEDAFNKSKQECYSLKCNLEKERMWTTKQLSQELESLEKVKRIKELEAIESRLE 226
QY  488 KAEKSLKDDLTCLKSFVMLVDERKNMMEKIKQEERKVDGLNKNFKVQCGKVMVDVTEKLI 547
Db  227 KTEFTLKEDLTCLKTLTVMFVDERKTMSEKLTCTEDKLQAASSQLQVEQNKVTVTEKLI 286
QY  548 EESKLLKSKSEMEKEYSLTKERDELGMGLRSEERSCELSVDDLKRLDGIIEEVER 607
Db  287 EETKRALKSKTDVEEKMYSVTKERDDLNKCLKAEBEKGNDLLSRVNMMLNQLQSLEAIEK 346
QY  608 EINRGR---SCKGSEFTCPEDNKRILETLTLEIERLKKRLLQQLQVEVGDLMKTEDEYDQLE 663
Db  347 DFLKNKLNQDSGKSTALHQENNKIKELSQEVERLKLKDKMKAIEDDLMKTEDEYETLE 406
QY  664 QKFRTEQDKANFLSQOLEEIKHQMAKHAIEKGEAVSQEAELRHRFRLEEAKSRLQAEV 723
Db  407 RRYANERDKAQFLSLEHVKMELAKYKLAEKTE-TSHEQWLFKRLQEEEAKSCHLSREV 465
QY  724 QALKEKIHLMNKKEDQLSQLQVDYSVLQQRFMEEETKNKNMGREVNLTKLELSKRYSR 783
Db  466 DALKEKIHGYMATEDLCHLQGDHSVCKKLNQQENNRDLGREIENLTKELEYRHFESK 525
QY  784 ALRPSGNRRMVDVPVASTGVQTEAVCGDAAEETP-----AVFIRKSFQE-ENH 832
Db  526 SLRPSLNGRRISRDPQFSKEVQTEAV-----DNEPFDYKSLIPLERAVINGOLYEESENQ 580
QY  833 IMSNLRQVGLKKMERSSVL-----DRYPPANELTMRKSWIPWMRKRE-----NGPSTPQ 883
Db  581 D-----EDPNDEGSVLSFKCSQSTPCPVN-----RKLWIPWMKSKEGHLQNG----- 622
QY  884 EKGPRPNQAGHPGELVLAPKQGPQLHIRVTPDHNSTATLEITSPTSE--EFFSSTTVI 941
Db  623 KMQTKPNANFVQPGDLVLSHTPGQPLHIKVTPDHQNTATLEITSPTTSPHYSYSTAVI 682
QY  942 PTLGNQKPRITIIPSPNVMSQPKK-SADPTLGPRAWSPVTITISREKSPGEGRSASFAD 1000
Db  683 PNCGTPKQRITILQNASITPVKSKTSTEDLMNLEQGMSPITMATFARAQTPESCGSLTPE 742
```

```

QY  1001 RPAS 1004
Db  743 RTMS 746

RESULT 13
Q86V48
ID  Q86V48 PRELIMINARY; PRT; 1026 AA.
AC  Q86V48;
DT  01-JUN-2003 (TrEMBLrel. 24, Created)
DT  01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  LUZP1 protein.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Testis;
RX  MEDLINE=22388257; PubMed=12477932;
RA  Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA  Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA  Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA  Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA  Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA  Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA  Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA  Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA  Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA  Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA  Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA  Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA  Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA  Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA  Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA  Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA  Jones S.J., Marra M.A.;
RT  "Generation and initial analysis of more than 15,000 full-length human
RL  and mouse cDNA sequences.";
RL  Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Testis;
RA  Strausberg R.;
RL  Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR  EMBL; BC051733; AAH51733.1; -.
DR  GO; GO:0005743; C:mitochondrial inner membrane; IEA.
DR  GO; GO:0005488; F:binding; IEA.
DR  GO; GO:0006810; P:transport; IEA.
DR  InterPro; IPR001993; Mitoch carrier.
DR  PROSITE; PS00215; MITOCH CARRIER; 1.
SQ  SEQUENCE 1026 AA; 114616 MW; A8A11163CB24DF5A CRC64;

Query Match      14.7%; Score 893; DB 4; Length 1026;
Best Local Similarity 31.9%; Pred. No. 1.5e-27;
Matches 283; Conservative 143; Mismatches 302; Indels 160; Gaps 32;

QY  323 QESHNRQLRLKLVGLSQRIEELEETNKSLOKAEELQELREKIAKGECSNLSLMAEVESL 382
Db  9 ETASSRHLRFKQLSRLRRLDELEATKNLQKAEDELDDLQDKVIAEGSNSSMLAEIVL 68
QY  383 RKRVLEMEGKDEEITKTEAQCRELKKKLQEEHHHSKELRLEVEKLOKRMSELEKLEAFS 442
Db  69 RQVRLRIEGKDEEIKRAEDLCRLMKKELEENLTRELKSEIERLQKRMAELEKLEAFS 128
QY  443 RSKSECTQLHLNLEKEKNLTKDLLNELEVVKSRVKELECSERLEKAEKSLKDDLTCLKS 502
Db  129 RSKNDCTQLCLSLNEERNLTKKISSLEMLRVKVKLESESSEDRLDKTEQSLASELEKLS 188
QY  503 FTVMLVDERKNMMEKIKQEERKVDGLNKNFKVQCGKVMVDVTEKLI EESKLLKSKSEMEE 562
Db  189 LTLSFVSERKYLNEKEKENEKLIKELTQ--KLEQNKKM-----NRDYTRNASNLER 237
```

QY	563	KEYSLTKERDEL	MGLRSEERS	CELSVDLLKKRLD	GIEEVEREI-NRGRSCKGSEFT	621
Db	238	NDLRI	---EDGISSTLPS	KESSR-----KGGLDYLKQVENETRNKSENEKNRN--	282	
QY	622	CPEDNKIRELT	LEIERLKKRLQQL	LEVVEGDLMKTEDEYDQLEQKFRTEQDKANFLSQOLE	681	
Db	283	-QEDNKVKDNL	QOEIEKLTQIKHFES	LEEBELKKMKSKNNDLQDNYLSEQNKNKLLASQLE	341	
QY	682	EIKHQMAKHA	KAIEKGEAVSQE	AELRHRFRLEEAKSRDLQAEVQALKEKIHEL--MNKEDQ	739	
Db	342	EIKLQIKKQK	ELENGEVEGEDA	FLSSKGRHERTKFRGHGSEASVSKHTARELSPQHKRER	401	
QY	740	LSQLQVDY	SVLQQRFMEEET	KNKMGREVLNLTKELELSKRYSRALRPSGNGRRMVDVPV	799	
Db	402	L-----	RNREFALNNEN-----YLSNR--QVSSPSFTNRR	---AKA	434	
QY	800	ASTGVQTEA	VGDAEEETPAV	FIRKSFQENHIMSNLRQVGLKKPMERSSVLDRYPPAA	859	
Db	435	SHMGVST	-----DSGTQET	-----KKT--EDRFVPSSSKSEG-KKSREQPSVLSRYPPAA	481	
QY	860	NELTMRKSW	IPWMRKRENGPST	-----POEKGPRPNQGAGHP	896	
Db	482	QEHS	--KAW-----KGTSKPGTESGLKGV	KEKTRTFTSDTTHG	SVSPDPLGRADKASDTS 534	
QY	897	GELVLAPK	QGOPL--HIRVTPD	HENSTATLEITSPTSEEF	FSSTTVIPTLGNQKPRITII 954	
Db	535	SETVFG	-KRGHVLNGSQVT-QAANS	GCSKAIGALASSRRSSS	---EGLSKGKKAANGL 588	
QY	955	PSPN	-----VMSQKPKSAD	-----PTLGP	RAMSPVTTITISREKSPEGG--- 994	
Db	589	EADNSCPN	SKAPVLSKYPYSCRSQENILQGF	STSHKEGVNQPAAV--VMEDSSP	HEALRC 646	
QY	995	---RSAFADR	PASPIQI---MTVST	SAAPTETIAVSPESQ	EVPMGRTILKVTPEKQT---- 1044	
Db	647	RVIKSSGRE	KPDSDDDLDIASLV	TAKLVNTTITPEPEPKQ	PNSREKAKTRGAPRTSLFE 706	
QY	1045	-----	VPAPVRKYN	SNANIITTEDNKI	HIHLGSQFK-RSPGPAAEGVSPVITVRPVNV 1096	
Db	707	NDKDAGME	NESVKSVRASNTMEL	PDTN-----GAGVKS	QRPFPSPREALRSRAIIKPVIV 761	
QY	1097	TAEKEV	-----STGTV	LRSPRNHLSSRPGASKVTSTITITP	VTTSSTR 1139	
Db	762	--DKDV	KKIMGGSGTETTLEKQ	KPVSKPGPNKVTSSITII	YPSDSSSPR 807	

RESULT 14

Q8TEH1	PRELIMINARY;	PRT; 1046 AA.
AC	Q8TEH1;	
DT	01-JUN-2002 (TrEMBLrel. 21, Created)	
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE	FLJ00226 protein (Fragment).	
GN	FLJ00226.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxId=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Spleen;	
RA	Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.;	
RT	"The nucleotide sequence of a long cDNA clone isolated from human spleen."	
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AK074153; BAB84979.1; -.	
DR	HSSP; P04268; IIC2.	
DR	Genew; HGNC:14985; IUZP1.	
DR	GO; GO:0005743; C:mitochondrial inner membrane; IEA.	
DR	GO; GO:0005488; F:binding; IEA.	
DR	GO; GO:0006810; P:transport; IEA.	
DR	InterPro; IPR001993; Mitoch_carrier.	

DR	PROSITE; PS00215; MITOCH_CARRIER; 1.			
FT	NON TER	1		
SQ	SEQUENCE	1046 AA; 116695 MW; 2942129B20048C7F CRC64;		
	Query Match	14.6%; Score 891; DB 4; Length 1046;		
	Best Local Similarity	31.9%; Pred. No. 1.8e-27;		
	Matches	283; Conservative 142; Mismatches 303; Indels 160; Gaps 32;		
QY	323 QESHNRQLRLKLVGLSQRIE	LEETNKSLOKAEEELQELREKIAKGECNSSLMAEVE	SL 382	
DB	2 ETASSRHLRFKLS	PRLDELEATKNLQKAEDLLDQDKVIAEGSNSSMLAEI	EVL 61	
QY	383 RKRVLMEGKDEEITKTEA	QCRLKKLOEEHHSHKELREVEKLQRMSELEKLEAF	S 442	
DB	62 RQVLRIEGKDEEIKRAED	LCRLMKEKLEEEENLTRELKSEIERLQKRMAELEKLEAF	S 121	
QY	443 RSKSECTQLHLNLEKEKN	LTKDLLNELEVVKSRVKELECSERLEKAELSLKDDLTK	LKS 502	
DB	122 RSKNDCTQLCLSLNEER	NLTKKISSELEMLRVKVKELSESSEDRDKTEQSLASE	LEKLS 181	
QY	503 FTVMLVDERKNNMEKI	KQERKVDGLNKNFKVQGVMDVTEKLIBESKLLKLGKSE	ME 562	
DB	182 LTLSFVSE	KYLVNEKEKENEKLIKELTQ--KLEQNKKM-----	NRDYTRNASNLER 230	
QY	563 KEYSLTKERDELMGKL	RSEERSCELSVDLLKKRLDGIIEVEREI--NRGRSCKG	SEFT 621	
DB	231 NDURI--	EDGISSTLPSKESRR-----KGGLDYLKQVENETRNK	SENEKNRN-- 275	
QY	622 CPEDNKIRELTLEIER	LKKRLQQLLEVVEGDLMKTEDEYDQLEQKFRTEQDKAN	FLSQOLE 681	
DB	276 -QEDNKVKDNLQOE	IEKLTQIKHFESLEEBELKKMKSKNNDLQDNYLSEQN	KNKLLASQLE 334	
QY	682 EIKHQMAKHAIEKGE	AVSQEALRHRFRLEEAKSRDLQAEVQALKEKIH	EL--MNKEDQ 739	
DB	335 EIKLQIKKQKELENGE	VEGEDAFLSSKGRHERTKFRGHGSEASVKHTARELS	PQHKRER 394	
QY	740 LSQLQVDYSVLQOR	FMEETKNKMGREVLNLTKELELSKRYSRALRPSGNGRR	MVDVPV 799	
DB	395 L-----	RNREFALNNEN-----YLSLNR--QVSSPSFTNRR	A--AKA 427	
QY	800 ASTGVQTEAVCGDA	EEETPAVFIRKSFQENHIMSNLRQVGLKKPMERSSVL	DRYPPAA 859	
DB	428 SHMGVST-----	DSGTQET---KKT--EDRFVPSSSKSEG-KKSREQPS	VLSRYPPAA 474	
QY	860 NELTMRKSWIPWM	RKRENGPST-----POEKGPRPNQGAGHP	896	
DB	475 QEHS--KAW-----	KGTSKPGTESGLKGVKEKTRTFTSDTTHG	SVSPDPLGRADKASDTS 527	
QY	897 GELVLAPKQGOPL--	HIRVTPDHENSTATLEITSPTSEFFSSTTVIPTLGNQ	KPRITII 954	
DB	528 SETVFG-KRGHVLNG	SQVT-QAANSGCCSKAIGALASSRRSSS---EGLSKGK	KAANGL 581	
QY	955 PSPN-----	VMSQKPKSAD-----PTLGP	RAMSPVTTITISREKSPEGG--- 994	
DB	582 EADNSCPN	SKAPVLSKYPYSCRSQENILQGFSTSHKEGVNQPAAV--	VMEDSSPHEALRC 639	
QY	995 ---	RSAFADRPASPIQI---MTVSTSAAPTEIAVSPESQ	EVPMGRTILKVTPEKQT--- 1044	
DB	640 RVIKSSGREKPD	DDDDLDIASLVTA	KLVTNTITPEPEPKQPN	SREKAKTRGAPRTSLFE 699
QY	1045 -----	VPAPVRKYN	SNANIITTEDNKI	HIHLGSQFK-RSPGPAAEGVSPVITVRPVNV 1096
DB	700 NDKDAGME	NESVKSVRASNTMEL	PDTN-----GAGVKS	QRPFPSPREALRSRAIIKPVIV 754
QY	1097 TAEKEV-----	STGTVLRS	PRNHLSSRPGASKVT	STITITPVTTSSTR 1139
DB	755 --DKDV	KKIMGGSGTET	LEKQKPVSKPGPNK	VTSSITIIYPSDSSSPR 800
RESULT	15			
Q99NG3				
ID	Q99NG3			
AC	Q99NG3;			
	PRELIMINARY;			
	PRT; 1067 AA.			

DT	01-JUN-2001 (TrEMBLrel. 17, Created)	
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)	
DE	Leucine zipper protein 1.	
GN	LUZpl.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxID=10090;	
RN	[1] _SEQUENCE FROM N.A.	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=ICR;	
RX	MEDLINE=96411647; PubMed=8812416;	
RA	Sun D.S., Chang A.C., Jenkins N.A., Gilbert D.J., Copeland N.G.,	
RA	Chang N.C.;	
RT	"Identification, molecular characterization, and chromosomal	
RT	localization of the cDNA encoding a novel leucine zipper motif-	
RT	containing protein.";	
RL	Genomics 36:54-62(1996).	
DR	EMBL; L49344; AAA98795.1; ..	
DR	HSSP; P04268; IIC2.	
DR	MGD; MGI:107629; Luzpl.	
SQ	SEQUENCE 1067 AA; 119289 MW; 381f1c16181749CA CRC64;	
Query Match 13.8%; Score 839; DB 11; Length 1067;		
Best Local Similarity 29.3%; Pred. No. 2e-25;		
Matches 275; Conservative 150; Mismatches 326; Indels 188; Gaps 28;		
QY	318 AKLANQE--SHNRQLRLKVLGLSQRIEELETNKSLOKAEELQELREKIAKGCNSSL 375	
Db	2 AELTNYKDAASNRLHRLFKLQSLRRLLDELEATKNLQRAEDELDDLQDKVIAEGSDST 61	
QY	376 MAEVESLRKRVLEMEGKDEEITKTEAQCRELKKLQEEHHHSKELRLEVEKQKRMSELE 435	
Db	62 LAEIEVLRQRVLKIEGDEEIKRAEDLCHTMKEKELEEEENLTRELKSEIERLQKRMVDLE 121	
QY	436 KLEEAFSRSKSECTQLHLNLEKEKNLTKDLLNELVVKSrvKELECSERLEKAELSLXD 495	
Db	122 KLEEAALSRSKNECSQLCLSLNEERNLTKKISSELMRLRVKVKKELESSEDRDKTEQSLVS 181	
QY	496 DLTKLKSFTVMLVDERKNMMEKIKQEEKVDGLNKNFKVEQGKVMVDTEKLIIEESKKLLK 555	
Db	182 ELEKLKSLTSLFVNERKYLNEKEKENEK-----IIKELTQKL-EQNKKM-- 224	
QY	556 LKSEMEEKEYSLTKERDELGMGLRSEERSCELSVDLLKKRLDGIIEVEREINRGRSC 615	
Db	225 --NRDHMRNASTFLERND---LRIEDGISSTLSSKESKRKGLDYLKQVENE-TRDKS- 276	
QY	616 KGSEFTCPEDNKIRELTLEIERLKKRLQQLVEVVEGLMKTEDYDQLEQKFRTEQDKANF 675	
Db	277 ENEKNRNQEDNKVKDLNQEIEKLTQIKHFESLEELKKMRAKNNDLQDNYLTELNRNS 336	
QY	676 LSQOLEEIKHQMAKHAIEKGEAVSQEAELRHRFLREAAKSRDLQAEVQALKEKIHLEL-- 733	
Db	337 LASQLEEIKLQVRKQKELNGDIEGEDAFLGRGRHERTKLKGHGSEASVSKHTRSRELSP 396	
QY	734 MNKEDQLSQLQVDYSLQQORFMEEEETKNKNMGREVLNLTKELELSKRYSRALRPSGNRR 793	
Db	397 QHKRERL-----RNREFALSNEHYSLSKQASSPVFTNKRAKA----- 435	
QY	794 MVDVPVASTGVQTEAVCGDAAEEETPAVFIRKSFOENHIMSNLRQVGLKPKMERSSVLD 853	
Db	436 -----SNMGMGTDs--GTQETKRTEDRFAPGSSHSEG-----KRGREQPSVLS 476	
QY	854 RYPPAANELTMRKSNIPWMRK-RENGPSTPQEKGRPNQGAGH---PGELV----- 900	
Db	477 RYPPAAQEHT--KVKGAPKPGTENGKLGKVEKTRTFSDSTHVSVPNDIVGKGDKTSDL 534	
QY	901 ----LAPKQGOPLHIRTVPDHNSTATLEITSPTSEEFFSSTTVIPTLGNQKPRITIIPS 956	
Db	535 SSEAHCGKRGQ-----VPGHA-SQGTQAVES-----SCSKAIGALSS----- 570	
QY	957 PNVMSQPKSADPTLGPERRAMSPVTITTISREKSPGGRSAFADRPASPI-----QIMTVS 1012	

Db	571 ----SQKASSEGLSGKKTANGLAADADFNSKAPILSKVPYSSRSQENILQGFSLPNKE 626	
QY	1013 TSAAPTEIAVSPESQEVPMGRITLK-----VTPEKQTV 1045	
Db	627 GVDQPVAVVMEDSSQHEALRCRVIKSSGREKPDSDDDLDIESFVTAKLVNTTITPEPEPK 686	
QY	1046 PAP-----VRKYNSNANIITTEDNKIHIHLGSGFKRSPGPAA----- 1082	
Db	687 PQPNSREKVKSRGGTTRTALFENDKNAAIENDSVKPTRPSSNAIEFPDANCAGVKQRPFS 746	
QY	1083 --EGVSPVITVRPVMVTAKEV-----STGTVLRSPRNHLSSRPGASKVTSTTITPVT 1135	
Db	747 PREALRSRAIIKPVI--DKDVKKIMGGSGTEVVLEKQKSTKSVTSKVTSSITITPSPDS 804	
QY	1136 SSTRGTQSVSGQDSSQRPTPTTRIPMSKGMKAGKPVVAA 1174	
Db	805 SGPRAVPS----EAPRERTST-----SNIQVGPPelta 834	

Search completed: September 7, 2004, 13:16:32
Job time : 160 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 7, 2004, 13:17:40 ; Search time 403 Seconds
(without alignments)
6119.292 Million cell updates/sec

Title: US-10-788-793-1
Perfect score: 7731
Sequence: 1 ccactgggtttcttcaaggga.....aaaaaaaaaaaaaaaaaaaa 4364

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues
Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10788793/runat_07092004_134020_10795/app_query.fasta_1.4551
-DB=A_Geneseq_29Jan04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10788793@cgn_1_1_642@runat_07092004_134020_10795 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_29Jan04:
1: geneseqp1980s:
2: geneseqp1990s:
3: geneseqp2000s:
4: geneseqp2001s:
5: geneseqp2002s:
6: geneseqp2003as:
7: geneseqp2003bs:
8: geneseqp2004s:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6082	78.7	1212	6 ABP97029	Abp97029 Rat L-FIL
2	5696.5	73.7	1213	4 AAM40016	Aam40016 Human pol
3	5696.5	73.7	1213	6 ABP97031	Abp97031 Human L-F
4	4838	62.6	965	6 ABP97030	Abp97030 Rat S-FIL
5	3051	39.5	653	4 AAB93085	Aab93085 Human pro
6	2586	33.4	551	4 AAB93244	Aab93244 Human pro
7	2452.5	31.7	1133	6 ABR62252	AbR62252 GPBP-inte
8	2444.5	31.6	1133	6 ABR62253	AbR62253 GPBP-inte
9	2444.5	31.6	1135	6 ABR62251	AbR62251 GPBP-inte
10	1926	24.9	893	6 ABR69651	AbR69651 Human CGD

11	1828	23.6	764	6 ABR62250	AbR62250 GPBP-inte
12	1675	21.7	785	6 ABR62257	AbR62257 GPBP-inte
13	1645.5	21.3	752	5 AAU84329	Aau84329 Protein D
14	1645.5	21.3	752	6 ABR92070	AbR92070 Human cer
15	1062.5	13.7	233	4 AAM41802	Aam41802 Human pol
16	929	12.0	365	4 AAB63771	Aab63771 Human pro
17	893	11.6	1053	6 ABR64181	AbR64181 Angiogene
18	893	11.6	1076	7 ADC31075	AdC31075 Human nov
19	893	11.6	1087	6 ABG74689	AbG74689 Human CGD
20	780.5	10.1	439	3 AAB58157	Aab58157 Lung canc
21	769.5	10.0	350	6 ABR62249	AbR62249 GPBP-inte
22	677	8.8	265	6 ABR62248	AbR62248 GPBP-inte
23	618.5	8.0	384	6 ABR62264	AbR62264 GPBP-inte
24	618.5	8.0	386	6 ABR62256	AbR62256 GPBP-inte
25	602	7.8	274	7 ADB65199	AdB65199 Human pro
26	569	7.4	190	4 AAB63826	Aab63826 Human pro
27	526.5	6.8	240	6 ABR62263	AbR62263 GPBP-inte
28	518.5	6.7	240	6 ABR62247	AbR62247 GPBP-inte
29	512	6.6	225	4 AAB63875	Aab63875 Human pro
30	476	6.2	2349	6 AAO16359	Aao16359 Human tra
31	476	6.2	2415	4 ABG20279	AbG20279 Novel hum
32	474	6.1	2400	4 ABG20278	AbG20278 Novel hum
33	472.5	6.1	2663	4 AAM39097	Aam39097 Human pol
34	470.5	6.1	2688	4 AAM40883	Aam40883 Human pol
35	469	6.1	1960	4 AAM78854	Aam78854 Human pro
36	469	6.1	1960	7 ADB70370	AdB70370 PAC 6802
37	469	6.1	1960	7 ADD47946	AdD47946 Human pro
38	469	6.1	2143	4 ABG01716	AbG01716 Novel hum
39	468.5	6.1	1959	5 ABJ10604	AbJ10604 Human nov
40	464.5	6.0	1790	6 ABR53116	AbR53116 Protein s
41	463.5	6.0	1963	4 AAM79838	Aam79838 Human pro
42	461.5	6.0	1881	5 ABP73809	Abp73809 Candida a
43	460.5	6.0	2633	4 ABG06505	AbG06505 Novel hum
44	454.5	5.9	1920	5 ABG97508	AbG97508 Human NOV
45	452	5.8	1857	5 AAU84350	Aau84350 Protein M

ALIGNMENTS

RESULT 1
ABP97029
ID ABP97029 standard; protein; 1212 AA.

XX ABP97029;
AC ABP97029;
DT 18-JUN-2003 (first entry)
XX
DE Rat L-FILIP protein SEQ ID NO:2.
XX
KW L-FILIP; S-FILIP; filamin-interacting protein; cell migration;
KW cell death; cytostatic; neuroprotective; immunosuppressive; cancer;
KW tumour metastasis; transplantation therapy.
XX
OS Rattus norvegicus.
XX
PN WO2003018804-A1.
XX
PD 06-MAR-2003.
XX
PF 29-JUL-2002; 2002WO-JP007676.
XX
PR 27-AUG-2001; 2001JP-00256910.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI Sato M, Nagano T;
XX
DR WPI; 2003-268423/26.
DR N-PSDB; ACC45354.
XX
PT Proteins controlling cell migration and cell death and their encoded
PT DNAs, applicable in developing drugs for treating or suppressing cancer
PT or tumor metastasis or as regulators of cell migration for

PT transplantation.

XX Claim 1; Page 54-60; 96pp; Japanese.

PS The present sequence represents rat L-FILIP which is a filamin-

XX interacting protein. L-FILIP has a function of controlling cell migration

CC and cell death. L-FILIP has cytotostatic, neuroprotective and

CC immunosuppressive activities. The L-FILIP protein can be used for

CC controlling cell migration and cell death, which is applicable in

CC developing drugs for treating or suppressing cancer or tumour metastasis

CC or as regulators of cell migration for transplantation therapy, and also

CC for controlling the mobility and cell death of nerve cells, promoting

CC decomposition of the actin-binding protein e.g. filamin-interacting

CC protein in the treatment of preinvenrtilcular nodular heterotopia

XX

SQ Sequence 1212 AA;

Alignment Scores:

Pred. No.:	0	Length:	1212
Score:	6082.00	Matches:	1212
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	78.67%	Indels:	0
DB:	6	Gaps:	0

US-10-788-793-1 (1-4364) x ABP97029 (1-1212)

QY	75	ATGAGATCACGAAATCAAGGTGAGAGAAAGTTTCATCTAACGGGCATGTCTCTGCCCCAAG	134
Db	1	MetArgSerArgAsnGlnGlyGluSerSerSerAsnGlyHisValSerCysProLys	20
QY	135	TCCTCCATCATFCAGCAGTGTGGTGAAGGGCCCCCTCAGAAGATGCRAAAAAAGAACAG	194
Db	21	SerSerIleIleSerSerAspGlyGlyLysGlyProSerGluAspAlaLysLysAsnLys	40
QY	195	GCCAAATCGAAGGAGGAGGATGTCTATGGCTCCGGAACTATCAAAAGGACCTCAAAACA	254
Db	41	AlaAsnArgLysGluGluaspValMetAlaSerGlyThrIleLysArgHisLeuLysPro	60
QY	255	TCTGGAGAAAGTGAAGAAAGACTAAGAACTCTGTGGAGTTATCCAAGGAGGACCTCATC	314
Db	61	SerGlyGluSerGluLysLysThrLysLysSerValGluLeuSerLysGluAspLeuIle	80
QY	315	CAGCTCCTGAGTATCATGGAAGGGGAGTTGCAGGCTCGAGAAGATGTCTATCCACATGCTG	374
Db	81	GlnLeuLeuSerIleMetGluGlyGluLeuGlnAlaArgGluaspValIleHisMetLeu	100
QY	375	AGGACAGAGAAAAACCAAGCCCGAGGTTCTGGAGGCACACTATGGATCTGCAGAACCTGAG	434
Db	101	ArgThrGluLysThrLysProGluValLeuGluAlaHisTyrGlySerAlaGluProGlu	120
QY	435	AAAGTGCTCGGTCCTGCACCGAGATGCCATCCTTGCTCAAGAGAAAGTCCATAGAGAA	494
Db	121	LysValLeuArgValLeuHisArgAspAlaIleLeuAlaGlnGluLysSerIleGlyGlu	140
QY	495	GACGTCTATGAGAAACCTTATCTCAGAGCTGGACAGACTGGAGGAAAGCAGAGGAGACG	554
Db	141	AspValTyrGluLysProIleSerGluLeuAspArgLeuGluGluLysGlnLysGluThr	160
QY	555	TACCGCGCATGCTAGAGCAGCTGCTGCTGGCTGAGAAGTGTCAACAGCGCACCCGTGTAC	614
Db	161	TyrArgArgMetLeuGluGlnLeuLeuAlaGluLysCysHisArgArgThrValTyr	180
QY	615	GAGCTGAGAACGAGAAGCAACACTGACTACTGAACAAGACGACGACTTCACC	674
Db	181	GluLeuGluAsnGluLysHisLysHisThrAspTyrMetAsnLysSerAspPheThr	200
QY	675	AACCTGCTGGAGCAGGACGAGAGGTTGAAAAAGCTCCTTGAACAAGAAAAAGCTTAC	734
Db	201	AsnLeuLeuGluGlnGluArgGluArgLeuLysLysLeuLeuGluGlnGluLysAlaTyr	220
QY	735	CAAGCCCGCAAAGAAAGGAAACCGCTAAGCGGTCAACAAACTTCGAGATGAGCTTGTG	794

Db	221	GlnAlaArgLysGluLysGluAsnAlaLysArgLeuAsnLysLeuArgAspGluLeuVal	240
QY	795	AAGCTCAAGTCCTTCGCCCTCATGTTGGTGGACGAGAGCGAGATGCACATCGACCAACTG	854
Db	241	LysLeuLysSerPheAlaLeuMetLeuValAspGluArgGlnMetHisIleGluGlnLeu	260
QY	855	GGCCTGCAGAGTCAGAAAGTCCAGGACCTCAGAAAGTCTCAGAAAGTGTAGGAGGAGAAAA	914
Db	261	GlyLeuGlnSerGlnLysValGlnAspLeuThrGlnLysLeuArgGluGluGluLys	280
QY	915	CTCAAAGCGGTCACTTACAAATCCAAAGAAAGACCCGCCAGAAGCTGCTCAAGTTAGAAGTG	974
Db	281	LeuLysAlaValThrTyrLysSerLysGluAspArgGlnLysLeuLeuLysLeuGluVal	300
QY	975	GACTTCGAACACAAAGGCTCGAGGTTTTCCCGAGGACGAAAGAGATGAACGCCAAATTG	1034
Db	301	AspPheGluHisLysAlaSerArgPheSerGlnGluHisGluGluMetAsnAlaLysLeu	320
QY	1035	GCGAATCAAGAATCTCACAACCGGCAACTTCGACTCAAACTGGTTGGCTTATCGCAAAGG	1094
Db	321	AlaAsnGlnGluSerHisAsnArgGlnLeuArgLeuLysLeuValGlyLeuSerGlnArg	340
QY	1095	ATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAGGCAGAGGAAGAGCTCCAGGAG	1154
Db	341	IleGluGluLeuGluGluThrAsnLysSerLeuGlnLysAlaGluGluGluGlnGlu	360
QY	1155	CTGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAACTCCAGTCTCATGGCGGAAGTGGAG	1214
Db	361	LeuArgGluLysIleAlaLysGlyGluCysGlyAsnSerSerLeuMetAlaGluValGlu	380
QY	1215	AGTCTGCGCAAGCGCTGTGAGATGGAGGGCAAGGATGAAGAGATCACGAAGACCGAG	1274
Db	381	SerLeuArgLysArgValLeuGluMetGluGlyLysAspGluGluIleThrLysThrGlu	400
QY	1275	GCCAGTCCCGGAGCTGAAGAAAGAGCTCCAAGAGGAAGAACACACAGCAAGGAACTT	1334
Db	401	AlaGlnCysArgGluLeuLysLysLysLeuGlnGluGluGluHisSerLysGluLeu	420
QY	1335	AGACTAGAAGTGGAGAAAGCTGCAGAAGAGGATGTCTGAGCTGGAGAAGCTGGAGGAGCG	1394
Db	421	ArgLeuGluValGluLysLeuGlnLysArgMetSerGluLeuGluLysLeuGluGluAla	440
QY	1395	TTAGCCGGAGTAAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAAGGAGAAGAC	1454
Db	441	PheSerArgSerLysSerGluCysThrGlnLeuHisLeuAsnLeuGluLysGluLysAsn	460
QY	1455	CTAACCAAAGACCTGTGAACCGAGCTGGAGGTGCTCAAGAGTCGAGTTAAAGAACTCGAA	1514
Db	461	LeuThrLysAspLeuLeuAsnGluLeuGluValValLysSerArgValLysGluLeuGlu	480
QY	1515	TGCTCCGAGAGTAGACTGGAGAAGCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTG	1574
Db	481	CysSerGluSerArgLeuGluLysAlaGluLeuSerLeuLysAspLeuThrLysLeu	500
QY	1575	AAGTCCTTCACTGTGATGCTGCTGGATGAGAGGAAAAATATGATGGAGAAAAATAAGCAA	1634
Db	501	LysSerPheThrValMetLeuValAspGluArgLysAsnMetMetGluLysIleLysGln	520
QY	1635	GAAGAGAGGAAAGTGGATGGGTTGAATAAAAACTTTAAGGTGGAGCAGGGAAAAAGTCATG	1694
Db	521	GluGluArgLysValAspGlyLeuAsnLysAsnPheLysValGluGlnGlyLysValMet	540
QY	1695	GATGTACGGAAAAAGCTAATCGAGGAAAGCAAGAAAGCTTTTAAAACTCAAATCTGAAATG	1754
Db	541	AspValThrGluLysLeuIleGluGluSerLysLysLeuLeuLysLeuLysSerGluMet	560
QY	1755	GAGGAAAAGGAGTACAGTCTGACAAAGGAGAGCGGATGAGCTGATGGGTAAACTGAGGAGC	1814
Db	561	GluGluLysGluTyrSerLeuThrLysGluArgAspGluLeuMetGlyLysLeuArgSer	580
QY	1815	GAAGAAGAAAGGTCCTGTGAACTGAGCTGCAGTGTAGACTTTACTAAAGAAAGCGGCTTGAT	1874
Db	581	GluGluGluArgSerCysGluLeuSerCysSerValAspLeuLeuLysLysArgLeuAsp	600

QY	1875	GGCATAGAGGAGGTAGAAAGGGAAATAAACCGAGGTAGGTCTGTGCAAGGGTCTGAGTTC	1934
Db	601	GlyIleGluGluValGluArgGluIleAsnArgGlyArgSerCysLysGlySerGluPhe	620
QY	1935	ACCTGCCCGGAAGACAATAAGATCAGAGAACTAACGCTTGAAATCGAGAGACTGAAGAA	1994
Db	621	ThrCysProGluAspAsnLysIleArgGluLeuThrLeuGluIleGluArgLeuLysLys	640
QY	1995	CGGCTCCAGCAGTTGAGGTTGGTGGAGGGGACTTGATGAAGACCGAGGACCAATATGAC	2054
Db	641	ArgLeuGlnGlnLeuGluValValGluGlyAspLeuMetLysThrGluAspGluTyrAsp	660
QY	2055	CAGTTGGAGCAGAAATTCAAGAACCGAGCAGGATAAGGCAAACTTCCTCTCCAGCAGCTC	2114
Db	661	GlnLeuGluGlnLysPheArgThrGluGlnAspLysAlaAsnPheLeuSerGlnGlnLeu	680
QY	2115	GAGGAAATCAACACCAAAATGGCCNAGCACAAAGCCATAGAGAAAGGGGAGCCGTGAGC	2174
Db	681	GluGluIleLysHisGlnMetAlaLysHisLysAlaIleGluLysGlyGluAlaValSer	700
QY	2175	CAGGAAGCCGAACCTCGACACAGGTTTCGGCTGGAGGAGGCTAAAAGTCGTGATTACAG	2234
Db	701	GlnGluAlaGluLeuArgHisArgPheArgLeuGluAlaLysSerArgAspLeuGln	720
QY	2235	GCCGAGGTGCAGGCTCTCAAGGAGAAGATCCAGAGCTGATGAACAAGGAGACCAAGCTG	2294
Db	721	AlaGluValGlnAlaLeuLysGluLysIleHisGluLeuMetAsnLysGluAspGlnLeu	740
QY	2295	TCTCAGTCCAAAGTCGACTATTTCGGTCCTTCAGCAAGATTATGGAAGAACTAAG	2354
Db	741	SerGlnLeuGlnValAspTyrSerValLeuGlnArgPheMetGluGluThrLys	760
QY	2355	AACAAGAACATGGGAGGGAGGTCTCTCAATCTGACCAAGGAGCTAGAGCTTCCAAGCGC	2414
Db	761	AsnLysAsnMetGlyArgGluValLeuAsnLeuThrLysGluLeuGluSerLysArg	780
QY	2415	TACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAAGGATGGTGGACGTGCTGGCC	2474
Db	781	TyrSerArgAlaLeuArgProSerGlyAsnGlyArgArgMetValAspValProValAla	800
QY	2475	TCCACTGGGGTGCAGACCGAGCGCGGTGTGCGGGGATGCTGCGGAGGAGGACCCCGCT	2534
Db	801	SerThrGlyValGlnThrGluAlaValCysGlyAspAlaAlaGluGluThrProAla	820
QY	2535	GTGTTTCATTCGCAATCCTTCCAGGAGGAAATCACATCATGAGTAATCTCGACAGGTA	2594
Db	821	ValPheIleArgLysSerPheGlnGluGluAsnHisIleMetSerAsnLeuArgGlnVal	840
QY	2595	GGCCTGAAGAAACCCATGGAACGGTCTCTCGTCTCGACAGGTATCCCCAGCAGCGAAT	2654
Db	841	GlyLeuLysLysProMetGluArgSerSerValLeuAspArgTyrProProAlaAlaAsn	860
QY	2655	GAGCTCACCATGAGGAAGTCTTGATTCCTTGATGAGAAAGAGAAACGGTCTCTCC	2714
Db	861	GluLeuThrMetArgLysSerTrpIleProTrpMetArgLysArgGluAsnGlyProSer	880
QY	2715	ACTCCGCAGGAGAAAGGGCCCGCCAGGCCAAACCCAGGGTGCAGGGCACCCCGGGAGCTGGTC	2774
Db	881	ThrProGlnGluLysGlyProArgProAsnGlnGlyAlaGlyHisProGlyGluLeuVal	900
QY	2775	CTAGCACCAAGCAGGGCCAGCCCTACACATCCGTGTGACACCAGATCATGAGAACAGC	2834
Db	901	LeuAlaProLysGlnGlyGlnProLeuHisIleArgValThrProAspHisGluAsnSer	920
QY	2835	ACTGCCACCTGGAGATCAAGCCCCACATCTGAAGAGTTTTTCTCTAGTACCACCGTC	2894
Db	921	ThrAlaThrLeuGluIleThrSerProThrSerGluGluPhePheSerSerThrThrVal	940
QY	2895	ATTCTTACCTTAGGCAACACAGAAACCAAGATAACCATTTATTCATCACCAATGTCTATG	2954
Db	941	IleProThrLeuGlyAsnGlnLysProArgIleThrIleIleProSerProAsnValMet	960

QY	2955	TCGCAAAAGCCCAAAAGTGCAGATCCTACTCTCGGGCCAGAACGAGCCATGTCCCCCTGTC	3014
Db	961	SerGlnLysProLysSerAlaAspProThrLeuGlyProGluArgAlaMetSerProVal	980
QY	3015	ACGATTACTACTATTTCCAGAGAGAAGAGCCCGGAAGGTGGAAGGAGCGCTTTGCCGAC	3074
Db	981	ThrIleThrThrIleSerArgGluLysSerProGluGlyArgSerAlaPheAlaAsp	1000
QY	3075	AGGCCTGCATCCCCCATCCAAATCATGACGGTGTCAACATCTGCAGCTCCCACTGAAATC	3134
Db	1001	ArgProAlaSerProIleGlnIleMetThrValSerThrSerAlaAlaProThrGluIle	1020
QY	3135	GCTGTCTCTCCTGAATCTCAGGAAGTGCCTATGGGAAGGACTATCCTCAAAGTCACCCCG	3194
Db	1021	AlaValSerProGluSerGlnGluValProMetGlyArgThrIleLeuLysValThrPro	1040
QY	3195	GAAAAACAAACTGTTCCAGCCCCCGTGGGAAGTACAACTCCAATGCTTAATATCATCACC	3254
Db	1041	GluLysGlnThrValProAlaProValArgLysTyrAsnSerAsnAlaAsnIleThr	1060
QY	3255	ACGGAAGACATAAAATTCACATTCACCTGGGTCTCAGTTTAAGGATCTCCTGGGCCT	3314
Db	1061	ThrGluAspAsnLysIleHisIleHisLeuGlySerGlnPheLysArgSerProGlyPro	1080
QY	3315	GCCGCTGAAGCGTGAGCCCAAGTTATCACCGTCCGGCCTGTCAACGTGACAGCGGAGAAG	3374
Db	1081	AlaAlaGluGlyValSerProValIleThrValArgProValAsnValThrAlaGluLys	1100
QY	3375	GAGGTTTCTACAGGCACAGTCTTCCGTCTCCAGGAACCCACCTCTCTTCAAGACCCGGT	3434
Db	1101	GluValSerThrGlyThrValLeuArgSerProArgAsnHisLeuSerSerArgProGly	1120
QY	3435	GCTAGCAAAAGTGACCAGCACTATAACTATAACCCCGGTCAACAACGTCTATCCACACGAG	3494
Db	1121	AlaSerLysValThrSerThrIleThrIleThrProValThrThrSerSerThrArgGly	1140
QY	3495	ACCCAATCAGTGTTCAGGACAAGATGGGTCTCTCAGCGGCCTACCCCCACCCGCTTCT	3554
Db	1141	ThrGlnSerValSerGlyGlnAspGlySerSerGlnArgProThrProThrArgIlePro	1160
QY	3555	ATGTCAAAAGGTATGAAAGCTGGAAAGCCAGTAGTGGCAGCCTCAGGAGCAGGAAATCTG	3614
Db	1161	MetSerLysGlyMetLysAlaGlyLysProValValAlaAlaSerGlyAlaGlyAsnLeu	1180
QY	3615	ACCAAAATCCAGCCTCGAGCTGAGACTCAGTCTATGAAAATAGAGCTGAAGAAATCTGCA	3674
Db	1181	ThrLysPheGlnProArgAlaGluThrGlnSerMetLysIleGluLeuLysLysSerAla	1200
QY	3675	GCCAGCAGCACTGCCTCTCTTGGAGGGGGGAAGGC	3710
Db	1201	AlaSerSerThrAlaSerLeuGlyGlyLysGly	1212
RESULT 2			
AAM40016			
ID	AAM40016 standard; protein; 1213 AA.		
XX	AAM40016;		
AC	AAM40016;		
XX	22-OCT-2001 (first entry)		
DT	Human polypeptide SEQ ID NO 3161.		
XX	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukaemia.		
XX	Homo sapiens.		
OS	XX		
XX	XX		
PN	WO200153312-A1.		
XX	XX		

Db 441 AlaPheSerLysSerLysSerGluCysThrGlnLeuHisLeuAsnLeuGluLysGluLys 460
QY 1452 AACCTAACCAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTC 1511
Db 461 AsnLeuThrLysAspLeuLeuAsnGluLeuGluValValLysSerArgValLysGluLeu 480
QY 1512 GAATGCTCCGAGAGTAGACTGGAGAAGGCCGAGTTAAGCCTCAAAAGATGACCTTACAAAG 1571
Db 481 GluCysSerGluSerArgLeuGluLysAlaGluLeuSerLeuLysAspAspLeuThrLys 500
QY 1572 CTGAAGTCTTCACTGTGATGCTGGTGATGAGAGGAAAAATATGATGGAGAAAAATAAG 1631
Db 501 LeuLysSerPheThrValMetLeuValAspGluValArgLysAsnMetMetGluLysIleLys 520
QY 1632 CAAGAAGAGAGGAAAGTGGATGGTTGAATAAAAACTTTAAGGTGGAGCAGGAAAAAGTC 1691
Db 521 GlnGluGluArgLysValAspGlyLeuAsnLysAsnPheLysValGluGlnGlyLysVal 540
QY 1692 ATGGATGTGACCGGAAAAAGCTAATCGAGGAAAGCAAGAAGCTTTTAAAACTCAAACTCTGAA 1751
Db 541 MetAspValThrGluLysLeuIleGluSerLysLysLeuLysLeuLysSerGlu 560
QY 1752 ATGGAGGAAAAAGGATACAGTCTGCAAAAGGAGAGGATGAGCTGATGGTAAACTGAGG 1811
Db 561 MetGluGluLysValTyrAsnLeuThrArgGluArgAspGluLeuIleGlyLysLeuLys 580
QY 1812 AGCGAAGAAGAAAGTCTGTGAAGTGAAGTGCAGTGCAGTGTAGACTTACTAAAGACGGCTT 1871
Db 581 SerGluGluGluLysSerSerGluLeuSerCysSerValAspLeuLeuLysLysArgLeu 600
QY 1872 GATGGCATAGAGGAGGTAGAAAGGGAATAAAACCGAGGTAGTTCGTGCAAGGGTCTGAG 1931
Db 601 AspGlyIleGluGluValGluArgGluIleThrArgGlyArgSerArgLysGlySerGlu 620
QY 1932 TTCACCTGCCCGGAAGACAATAAGATCAGAGAACTAAACGCTTGAAATCGAGAGACTGAAG 1991
Db 621 LeuThrCysProGluAspAsnLysIleLysGluLeuThrLeuGluIleGluArgLeuLys 640
QY 1992 AAACGGCTCCAGCAGTTGGAGGTGGTGGAGGGGACTTGATGAAGACCGGACGAATAT 2051
Db 641 LysArgLeuGlnGlnLeuGluValValGluGluValGluGluMetLysThrGluAspGluTyr 660
QY 2052 GACCAGTTGGAGCAGAAGTTCAGAACCGAGCAGGATAAGGCAAACTTCCTCTCCCGAGCAG 2111
Db 661 AspGlnLeuGluGlnLysPheArgThrGluGlnAspLysAlaAsnPheLeuSerGlnGln 680
QY 2112 CTCGAGGAAATCAAAACACCAAAATGGCCAAAGCACAAAGCCATAGAGAAAGGGAGCCGTG 2171
Db 681 LeuGluGluIleLysHisGlnIleAlaLysAsnLysAlaIleGluLysGlyGluValVal 700
QY 2172 AGCCAGGAAGCCGAACTGCGACACAGGTTTCGGCTGGAGGAGGCTTAAAGTCGTGATTTA 2231
Db 701 SerGlnGluAlaGluLeuArgHisArgPheArgLeuGluAlaLysSerArgAspLeu 720
QY 2232 CAGGCCGAGGTGACGCTCTCAAGGAGAAGATCCACGAGCTGATGAACAAGGAAGACCCAG 2291
Db 721 LysAlaGluValGlnAlaLeuLysGluLysIleHisGluLeuMetAsnLysGluAspGln 740
QY 2292 CTGTCTCAGCTCCAAGTCGACTATTTCGGTCTTCAGCAAAAGATTTATGGAGAAGAAACT 2351
Db 741 LeuSerGlnLeuGlnValAspTyrSerValLeuGlnArgPheMetGluGluAsn 760
QY 2352 AAGAACAAAGACATGGGAGGGAGGTCTCAATCTGACCAAGGAGCTAGAGCTTTCGAAG 2411
Db 761 LysAsnLysAsnMetGlyGlnGluValLeuAsnLeuThrLysGluLeuGluLeuSerLys 780
QY 2412 CGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAAGGATGGTGGACGTGCCTGTG 2471
Db 781 ArgTyrSerArgAlaLeuArgProSerValAsnGlyArgArgMetValAspValProVal 800
QY 2472 GCCTCCACTGGGTGCAGACCGAGGCGGTGTGCGGGGATGCTGCGGAGGAGGAGACCCCG 2531
Db 801 ThrSerThrGlyValGlnThrAspAlaValSerGlyGluAlaAlaGluGluGluThrPro 820

QY 2532 GCTGTGTTTCATTCGCAAAATCCTTCCAGGAGGAAAAATCACATCATGAGTAATCTTCGACAG 2591
Db 821 AlaValPheIleArgLysSerPheGlnGlnGluAsnHisIleMetSerAsnLeuArgGln 840
QY 2592 GTAGGCCCTGAAGAAACCCATGGAACCGTCTCGGTCTCTCGACAGGTATCCCCCAGCAGCG 2651
Db 841 ValGlyLeuLysLysProValGluArgSerSerValLeuAspArgTyrProProAlaAla 860
QY 2652 AATGAGCTCACCATGAGGAAGTCTTGGATTCCTTGGATGAGAAAAAGAGAAAAACGGTCCT 2711
Db 861 AsnGluLeuThrMetArgLysSerTrpIleProTrpMetArgLysArgGluAsnGlyPro 880
QY 2712 TCCACTCCGAGGAGAAAGGGCCCGAGCCCAACAGGGTGCAGGGCACCCCGGGGAGCTG 2771
Db 881 SerIleThrGlnGluLysGlyProArgThrAsnSerSerProGlyHisProGlyGluVal 900
QY 2772 GTCCTAGCACCAAGAGCAGGGCCAGCCCTACACATCCGTGTGACACCATCATGAGAAC 2831
Db 901 ValLeuSerProLysGlnGlyGlnProLeuHisIleArgValThrProAspHisGluAsn 920
QY 2832 AGCACTGCCACCTGGAGATCACAAAGCCACACATCTGAAGAGTTTTTCTCTAGTACCACC 2891
Db 921 SerThrAlaThrLeuGluIleThrSerProThrSerGluGluPhePheSerSerThrThr 940
QY 2892 GTCATTCTCTACCTTAGGCAACCAAGAAACCAAGAATAACCATTTATTCATCACCCCAATGTC 2951
Db 941 ValIleProThrLeuGlyAsnGlnLysProArgIleThrIleIleProSerProAsnVal 960
QY 2952 ATGTCGCAAAAGCCCCAAAAGTGCAGATCTTACTCTCGGCCAGAACGAGCCATGTCCCCCT 3011
Db 961 MetProGlnLysGlnLysSerGlyAspThrThrLeuGlyProGluArgAlaMetSerPro 980
QY 3012 GTCACGATTACTACTATTTCAGAGAGAAGAGCCCGGAAGGTGGAAGGAGCGCTTTGCC 3071
Db 981 ValThrIleThrThrPheSerArgGluLysThrProGluSerGlyArgGlyAlaPheAla 1000
QY 3072 GACAGGCTGCATCCCCCATCCAAATCATGACGGTGTCAACATCTGCAGCTCCCACTGAA 3131
Db 1001 AspArgProThrSerProIleGlnIleMetThrValSerThrSerAlaAlaProAlaGlu 1020
QY 3132 ATCGCTGTCTCTCTGAATCTCAGGAAGTGCCTATGGGAAGGACTATCTCTCAAGTCACC 3191
Db 1021 IleAlaValSerProGluSerGlnGluMetProMetGlyArgThrIleLeuLysValThr 1040
QY 3192 CCGGAAAAACAACCTGTTCCAGCCCCCGTCGCGGAAGTACAACCTCCAATGCTAATATCATC 3251
Db 1041 ProGluLysGlnThrValProThrProValArgLysTyrAsnSerAsnAlaAsnIleIle 1060
QY 3252 ACCACGGAAACAATAAAATTCACATTCACCTGGGTCTCAGTTTAAGCGATCTCCTGGG 3311
Db 1061 ThrThrGluAspAsnLysIleHisIleHisLeuGlySerGlnPheLysArgSerProGly 1080
QY 3312 CCTGCCGCTGAAGCGCTGAGCCCCAGTTATCACCGTCCGGCCTGTCAACGTGACAGCGGAG 3371
Db 1081 ThrSerGlyGluGlyValSerProValIleThrValArgProValAsnValThrAlaGlu 1100
QY 3372 AAGGAGGTTTCTACAGGCACAGTCCCTTCGCTCTCCCAGGAACCACTCTCTTCAAGACCC 3431
Db 1101 LysGluValSerThrGlyThrValLeuArgSerProArgAsnHisLeuSerSerArgPro 1120
QY 3432 GGTGTAGCAAGTGACCAGCACTATACTATAACCCCGGTCAACACGTATCCACACGGA 3491
Db 1121 GlyAlaSerLysValThrSerThrIleThrIleThrProValThrThrSerSerAlaArg 1140
QY 3492 GGAACCCCAATCAGTGTGAGCAAGATGGGTTCATCTCAGCGGCCTACCCCGACCCGCATT 3551
Db 1141 GlyThrGlnSerValSerGlyGlnAspGlySerSerGlnArgProThrProThrArgIle 1160
QY 3552 CCTATGTCAAAAGGTATGAAAGCTGGAAGCCAGTAGTGGCAGCCTCAGGAGCAGGAAT 3611
Db 1161 ProMetSerLysGlyMetLysAlaGlyLysProValValAlaAlaProGlyAlaGlyAsn 1180

QY 3612 CTGACCAAAATCCAGCCTCGAGCTGAGACTCAGTCTATGAAATAGAGCTGAAGAAATCT 3671
Db LeuThrLysPheGluProArgAlaGluThrGlnSerMetLysIleGluLeuLysLysSer 1200

QY 3672 GCAGCCAGCAGCACTGCCTCTCTTGGAGGGGGGAAGGGC 3710
Db AlaAlaSerSerThrThrSerLeuGlyGlyGlyLysGly 1213

RESULT 3
ID ABP97031 standard; protein; 1213 AA.
XX ABP97031;
AC
XX 18-JUN-2003 (first entry)
DT
XX Human L-FILIP protein SEQ ID NO:6.
DE
XX L-FILIP; S-FILIP; filamin-interacting protein; cell migration;
KW cell death; cytostatic; neuroprotective; immunosuppressive; cancer;
KW tumour metastasis; transplantation therapy.
XX Homo sapiens.

OS
XX WO2003018804-A1.
XX 06-MAR-2003.
XX

PF 29-JUL-2002; 2002WO-JP007676.
XX
PR 27-AUG-2001; 2001JP-00256910.
XX

XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.

PA Sato M, Nagano T;
XX
PI WPI; 2003-268423/26.
DR N-PSDB; ACC45356.
XX

XX Proteins controlling cell migration and cell death and their encoded
PT DNAs, applicable in developing drugs for treating or suppressing cancer
PT or tumor metastasis or as regulators of cell migration for
PT transplantation.

PS Claim 7; Page 82-88; 96pp; Japanese.

XX The present sequence represents human L-FILIP which is a filamin-
CC interacting protein. L-FILIP has a function of controlling cell migration
CC and cell death. L-FILIP has cytostatic, neuroprotective and
CC immunosuppressive activities. The L-FILIP protein can be used for
CC controlling cell migration and cell death, which is applicable in
CC developing drugs for treating or suppressing cancer or tumour metastasis
CC or as regulators of cell migration for transplantation therapy, and also
CC for controlling the mobility and cell death of nerve cells, promoting
CC decomposition of the actin-binding protein e.g. filamin-interacting
CC protein in the treatment of preiventrilcular nodular heterotopia

XX SQ Sequence 1213 AA;

Alignment Scores:
Pred. No.: 0 Length: 1213
Score: 5696.50 Matches: 1134
Percent Similarity: 96.54% Conservative: 37
Best Local Similarity: 93.49% Mismatches: 41
Query Match: 73.68% Indels: 1
DB: 6 Gaps: 1

US-10-788-793-1 (1-4364) x ABP97031 (1-1213)

QY 75 ATGAGATCAGGAAATCAAGGTGGAGAAAGTTTCATCTAACGGGCATGCTCTCGCCCAAG 134
Db 1 MetArgSerArgAsnGlnGlyGlyGluSerAlaSerAspGlyHisIleSerCysProLys 20

QY 135 TCCTCCATCATCAGCAGTGATGGTGAAGGGCCCTCAGAAGATGCA---AAAAAGAAC 191
Db ProSerIleIleGlyAsnAlaGlyGluLysSerLeuSerGluAspAlaLysLysLys 40

QY 192 AAGGCCAATCGGAAGGAGGAGGATGTCATGGCTTCCGGAACATATCAAAGGCACCTCAA 251
Db LysSerAsnArgLysGluAspValMetAlaSerGlyThrValLysArgHisLeuLys 60

QY 252 CCATCTGGAGAAAAGTGAGAAAAAGACTAAGAAAGTCTGTGGAGTTATCCAGGAGGACCTC 311
Db ThrSerGlyGluCysGluArgLysThrLysLysSerLeuGluLeuSerLysGluAspLeu 80

QY 312 ATCCAGCTCCTGAGTATCATGGAAGGGAGTTGCAGGTCGAGAAGATGTCATCCACATG 371
Db IleGlnLeuLeuSerIleMetGluGlyGluLeuGlnAlaArgGluAspValIleHisMet 100

QY 372 CTGAGGACAGAGAAAACCAAGCCCGAGGTCTTGGAGGCACACTATGGATCTGCAGAACCT 431
Db LeuLysThrGluLysThrLysProGluValLeuGluAlaHisTyrGlySerAlaGluPro 120

QY 432 GAGAACTGCTTCGGGTCTGCACCGAGATGCCATCCTTGCTCAAGAGAAAGTCCATAGGA 491
Db GluLysValLeuArgValLeuHisArgAspAlaIleLeuAlaGlnGluLysSerIleGly 140

QY 492 GAAGACGCTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAAAAAGCAGAGGAG 551
Db GluAspValTyrGluLysProIleSerGluLeuAspArgLeuGluGluLysGlnLysGlu 160

QY 552 ACGTACCGCCGCATGCTAGAGCAGCTGCTGTGGTGAGAAAGTGTCAAGGGCGCACCGTG 611
Db ThrTyrArgArgMetLeuGlnLeuLeuAlaGluLysCysHisArgArgThrVal 180

QY 612 TACGAGCTGGAGAACGAGAAGCACAAGCACACTGACTACATGAACAAGAGCGACGACTTC 671
Db TyrGluLeuGluAsnGluLysHisLysHisThrAspTyrMetAsnLysSerAspAspPhe 200

QY 672 ACCAACCTGCTGGACGAGCGAGAGAGGTTTGAATAAGCTCCTTGAACAAGAAAAAGCT 731
Db ThrAsnLeuLeuGluGlnGluArgGluArgLeuLysLysLeuLeuGlnGluLysAla 220

QY 732 TACCAAGCCCGCAAAGAAAACGCTAAGCGGCTCAACAACCTTCGAGATGAGCTT 791
Db TyrGlnAlaArgLysGluLysGluAsnAlaLysArgLeuAsnLysLeuArgAspGluLeu 240

QY 792 GTGAAGCTCAAGTCTTCGCCCTCATGTTGGTGACGAGGCGCAGATGCACATCGAGCAA 851
Db ValLysLeuLysSerPheAlaLeuMetLeuValAspGluArgGlnMetHisIleGluGln 260

QY 852 CTGGCCCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAGCTGAGGGAGGAAGAA 911
Db LeuGlyLeuGlnSerGlnLysValGlnAspLeuThrGlnLysLeuArgGluGluGlu 280

QY 912 AAACCTCAAAGCGGTCACTTACAAATCCAAGGAAGACCCGCCAGAAGCTGCTCAAGTTAGAA 971
Db LysLeuLysAlaIleThrSerLysSerLysGluAspArgGlnLysLeuLeuLysLeuGlu 300

QY 972 GTGGACTTCGAACACAAAGGCCTCGAGGTTTTCCAGGAGCACCGAAGAGATGAACGCCAAA 1031
Db ValAspPheGluHisLysAlaSerArgPheSerGlnGluHisGluGluMetAsnAlaLys 320

QY 1032 TTGGCGAATCAAGAATCTCACACCGGCAACTTCGACTCAAACTGGTTGGCTTATCGCAA 1091
Db LeuAlaAsnGlnGluSerHisAsnArgGlnLeuArgLeuLysLeuValGlyLeuThrGln 340

QY 1092 AGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAAGCAGAGGAAGAGCTCCAG 1151
Db ArgIleGluGluLeuGluGluThrAsnLysAsnLeuGlnLysAlaGluGluLeuGln 360

QY 1152 GAGCTGAGAGAGAAAATTGCCAAAGGGGAATGTGGAAACTCAGTCTCATGGCGGAAGTG 1211
Db GluLeuArgAspLysIleAlaLysGlyGluCysGlyAsnSerSerLeuMetAlaGluVal 380

QY 1212 GAGAGTCTGCGCAAGCGCGTGCTTGAGATGCGGGGCAAGGATGAAGAGATCACGAAGACC 1271

Db 381 GluAsnLeuArgLysArgValLeuGluMetGluGlyLysAspGluGluIleThrLysThr 400
QY 1272 GAGGCCAGTGCCGGAGCTGAAGAAGACTCCAAGAGGAAGAACAACACAGCAAGGAA 1331
Db 401 GluSerGlnCysArgGluLeuArgLysLysLeuGlnGluGluGluHisSerLysGlu 420
QY 1332 CTTAGACTAGAAGTGGAGAAGCTGCAGAAGAGGTGTCTGAGCTGGAGAAGCTGGAGGAA 1391
Db 421 LeuArgLeuGluValGluLysLeuGlnLysArgMetSerGluLeuGluLysLeuGlu 440
QY 1392 GCGTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAGGAGAAG 1451
Db 441 AlaPheSerLysSerLysSerGluCysThrGlnLeuHisLeuAsnLeuGluLysGlu 460
QY 1452 AACCTAACCAAGACTGCTGAACGAGCTGGAGTGGTCAAGAGTCGAGTTAAAGAACTC 1511
Db 461 AsnLeuThrLysAspLeuLeuAsnGluLeuGluValValLysSerArgValLysGluLeu 480
QY 1512 GAATGCTCCGAGACTAGACTGGAGAAGCCGAGTTAAGCCTCAAAGATGACCTTACAAAG 1571
Db 481 GluCysSerGluSerArgLeuGluLysAlaGluLeuSerLeuLysAspAspLeuThrLys 500
QY 1572 CTGAAGTCCTTCACTGTGATGCTGGTGGATGAGAGGAAAAATATGATGGAGAAAAATAAG 1631
Db 501 LeuLysSerPheThrValMetLeuValAspGluArgLysAsnMetMetGluLysIleLys 520
QY 1632 CAAGAAGAGAGGAAGTGGATGGGTTGAATAAAAACTTTAAGGTGGAGCAGGAAAAAGTC 1691
Db 521 GlnGluGluArgLysValAspGlyLeuAsnLysAsnPheLysValGlnGlyLysVal 540
QY 1692 ATGGATGTACCGGAAAAGCTAATCGAGGAAAGCAAGAAGCTTTTAAAACTCAAATCTGAA 1751
Db 541 MetAspValThrGluLysLeuIleGluSerLysLysLeuLysLeuLysSerGlu 560
QY 1752 ATGGAGGAAAAGGAGTACAGTCTGACAAAAGGAGGAGTATGAGCTGATGGTAAACTGAGG 1811
Db 561 MetGluGluLysValTyrAsnLeuThrArgGluArgAspGluLeuIleGlyLysLeuLys 580
QY 1812 AGCGAAGAAGAAAGTCTGTGAACGTGAGCTGCAGTGTAGACTTACTAAAGAAGCGGCTT 1871
Db 581 SerGluGluGluLysSerSerGluLeuSerCysSerValAspLeuLeuLysLysArgLeu 600
QY 1872 GATGGCATAGAGGAGGTAGAAAGGGAATAAACCGAGGTAGGTGCTGTCAGGGGTCTGAG 1931
Db 601 AspGlyIleGluGluValGluArgGluIleThrArgGlyArgSerArgLysGlySerGlu 620
QY 1932 TTCACCTGCCGGAAGACAATAAGATCAGAGAACTAACGCTTGAAATCGAGAGACTGAAG 1991
Db 621 LeuThrCysProGluAspAsnLysIleLysGluLeuThrLeuGluIleGluArgLeuLys 640
QY 1992 AAACGGCTCCAGCAGTGGAGGTGGTGGAGGGGACTTGATGAAGACCCGAGGACGAATAT 2051
Db 641 LysArgLeuGlnGlnLeuGluValValGluGlyAspLeuMetLysThrGluAspGluTyr 660
QY 2052 GACCAGTTGGAGCAGAAGTTCAGAACCGAGCAGGATAAGGCAAACTTCCTCTCCAGCAG 2111
Db 661 AspGlnLeuGluGlnLysPheArgThrGluGlnAspLysAlaAsnPheLeuSerGlnGln 680
QY 2112 CTCGAGGAATCAAACACCAATGGCCAAGCACAAAGCCATAGAGAAAGGGAGGCCGTG 2171
Db 681 LeuGluGluIleLysHisGlnIleAlaLysAsnLysAlaIleGluLysGlyGluValVal 700
QY 2172 AGCCAGGAAGCCGAACACTCGCACACAGGTTTCGGTGGAGGAGGCTAAAGTCGTGATTTA 2231
Db 701 SerGlnGluAlaGluLeuArgHisArgPheArgLeuGluGluAlaLysSerArgAspLeu 720
QY 2232 CAGGCCGAGGTGCAGGCTCTCAAGGAGAAGATCCACGAGCTGATGAACAAGGAAGACCAG 2291
Db 721 LysAlaGluValGlnAlaLeuLysGluLysIleHisGluLeuMetAsnLysGluAspGln 740
QY 2292 CTGTCTCAGCTCCAAGTCGACTATTTCGGTCTCTCAGCAAAGATTTATGGAAGAAGAACT 2351

Db 741 LeuSerGlnLeuGlnValAspTyrSerValLeuGlnArgPheMetGluGluGluAsn 760
QY 2352 AAGAACAAGAACATGGGGAGGAGGTCTCAATCTGACCAAGGAGCTAGAGCTTTCCAAG 2411
Db 761 LysAsnLysAsnMetGlyGlnGluValLeuAsnLeuThrLysGluLeuGluLeuSerLys 780
QY 2412 CGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAAGGATGGTGGAGCTGCTGTG 2471
Db 781 ArgTyrSerArgAlaLeuArgProSerValAsnGlyArgArgMetValAspValProVal 800
QY 2472 GCCTCCACTGGGTGCAGACCCGAGCGGTGTCTCGGGGATGCTGCGGAGGAGAGACCCCG 2531
Db 801 ThrSerThrGlyValGlnThrAspAlaValSerGlyGluAlaAlaGluGluThrPro 820
QY 2532 GCTGTGTTTCATTCCGCAAAATCCTTCCAGGAGGAAAAATCACATCATGATGATTAATCTTCGACAG 2591
Db 821 AlaValPheIleArgLysSerPheGlnGluGluAsnHisIleMetSerAsnLeuArgGln 840
QY 2592 GTAGGCCCTGAAGAAAACCCATGGAACCGTCTCTCGGTCTCTCGACAGGTATCCCCCAGCAGCG 2651
Db 841 ValGlyLeuLysLysProValGluArgSerSerValLeuAspArgTyrProProAlaAla 860
QY 2652 AATGAGCTCACCATGAGGAAGTCTTGGATTCTCTGGATGAGAAAAAGAAAAACGGTCTCT 2711
Db 861 AsnGluLeuThrMetArgLysSerTrpIleProTrpMetArgLysArgGluAsnGlyPro 880
QY 2712 TCCACTCCGCAGGAGAAAGGGCCAGGCCCAACAGGGTGCAGGGCACCCCGGGGAGCTG 2771
Db 881 SerIleThrGlnGluLysGlyProArgThrAsnSerSerProGlyHisProGlyGluVal 900
QY 2772 GTCCTAGCACCAAGAGCGGGCCAGCCCTTACACATCCGTGTGACACCCAGATCATGAGAAC 2831
Db 901 ValLeuSerProLysGlnGlyGlnProIleHisIleArgValThrProaspHisGluAsn 920
QY 2832 AGCACTGCCACCCCTGGAGATCACAAAGCCCCACATCTGAAGAGTTTTCTCTAGTACCACC 2891
Db 921 SerThrAlaThrLeuGluIleThrSerProThrSerGluGluPhePheSerSerThrThr 940
QY 2892 GTCATTCTTACCTTAGGCAACCAGAAACCAAGAAATAACCATTTATTCATCACCCAATGTC 2951
Db 941 ValIleProThrLeuGlyAsnGlnLysProArgIleThrIleIleProSerProAsnVal 960
QY 2952 ATGTCGCAAAAGCCCAAAAGTGCAGATCCTACTCTCGGCCCAGAACCGAGCATGTCCCT 3011
Db 961 MetProGlnLysGlnLysSerGlyAspThrThrLeuGlyProGluArgAlaMetSerPro 980
QY 3012 GTCACGATTACTACTATTTTCCAGAGAGAGAGCCCGGAAGGTGGAAGGAGCGCTTTGCC 3071
Db 981 ValThrIleThrThrPheSerArgGluLysThrProGluSerGlyArgGlyAlaPheAla 1000
QY 3072 GACAGGCCCTGCATCCCCCATCCAAATCATGACGGTGTCAACATCTGCAGCTCCCCTGAA 3131
Db 1001 AspArgProThrSerProIleGlnIleMetThrValSerThrSerAlaAlaProAlaGlu 1020
QY 3132 ATCGTGTCTCTCCTGAATCTCAGGAAGTCCTATGGGAAGGACTATCTCTCAAAGTCACC 3191
Db 1021 IleAlaValSerProGluSerGlnGluMetProMetGlyArgThrIleLeuLysValThr 1040
QY 3192 CCGGAAAAACAACACTGTTCCAGCCCTCCAGCCCTCCGGCTGTCAACGTGACAGCGGAG 3251
Db 1041 ProGluLysGlnThrValProThrProValArgLysTyrAsnSerAsnAlaAsnIleIle 1060
QY 3252 ACCACGGAGACAATAAAATTCACATTCACCTGGGTTCTCAGTTTAAAGCGATCTCTGGG 3311
Db 1061 ThrThrGluAspAsnLysIleHisIleHisLeuGlySerGlnPheLysArgSerProGly 1080
QY 3312 CCTGCCGCTGAAGCGGTGAGCCCCAGTTATCACCGTCCGGCTGTCAACGTGACAGCGGAG 3371
Db 1081 ThrSerGlyGluGlyValSerProValIleThrValArgProValAsnValThrAlaGlu 1100
QY 3372 AAGGAGGTTTCTACAGGCACAGTCTCTCGCTCTCCAGGAAACCACTCTCTTCAAGACCC 3431
Db 1101 LysGluValSerThrGlyThrValLeuArgSerProArgAsnHisLeuSerSerArgPro 1120

QY 3432 GGTGCTAGCAAGTGAACGACACTATAACTATAAACCCCGGTCAACAACGTCTATCCACACGA 3491
|||||
Db 1121 GlyAlaSerLysValThrSerThrIleThrIleThrProValThrThrSerSerAlaArg 1140
|||||
QY 3492 GGAACCCAAATCAGTGTGAGACAAGATGGGTCAATCTCAGCGGCCTACCCACCCGCATT 3551
|||||
Db 1141 GlyThrGlnSerValSerGlyGlnAspGlySerSerGlnArgProThrProThrArgIle 1160
|||||
QY 3552 CCTATGTCAAAGGTATGAAAGCTGGAAGCCAGTAGTGGCAGCCTCAGGAGCAGGAAAT 3611
|||||
Db 1161 ProMetSerLysGlyMetLysAlaGlyLysProValValAlaAlaProGlyAlaGlyAsn 1180
|||||
QY 3612 CTGACCAAAATCCAGCCTCGAGCTGAGACTCAGTCTATGAAAAATAGAGCTGAAGAAATCT 3671
|||||
Db 1181 LeuThrLysPheGluProArgAlaGluThrGlnSerMetLysIleGluLeuLysLysSer 1200
|||||
QY 3672 GCAGCCAGCAGCACTGCCTCTCTTTGGAGGGGGGAAGGC 3710
|||||
Db 1201 AlaAlaSerSerThrThrSerLeuGlyGlyGlyLysGly 1213
|||||

RESULT 4
ABP97030
ID ABP97030 standard; protein; 965 AA.
XX
AC ABP97030;
XX
DT 18-JUN-2003 (first entry)
XX
DE Rat S-FILIP protein SEQ ID NO:4.
XX
KW L-FILIP; S-FILIP; filamin-interacting protein; cell migration;
KW cell death; cytosstatic; neuroprotective; immunosuppressive; cancer;
KW tumour metastasis; transplantation therapy.
OS Rattus norvegicus.
XX
PN WO2003018804-A1.
XX
PD 06-MAR-2003.
XX
PF 29-JUL-2002; 2002WO-JP007676.
XX
PR 27-AUG-2001; 2001JP-00256910.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI Sato M, Nagano T;
XX
DR WPI; 2003-268423/26.
DR N-PSDB; ACC45355.
XX
PT Proteins controlling cell migration and cell death and their encoded
PT DNAs, applicable in developing drugs for treating or suppressing cancer
PT or tumor metastasis or as regulators of cell migration for
PT transplantation.
XX

Claim 4; Page 68-73; 96pp; Japanese.
XX
XX The present sequence represents rat S-FILIP which is a filamin-
CC interacting protein. S-FILIP has a function of controlling cell migration
CC and cell death. S-FILIP has cytosstatic, neuroprotective and
CC immunosuppressive activities. The S-FILIP protein can be used for
CC controlling cell migration and cell death, which is applicable in
CC developing drugs for treating or suppressing cancer or tumour metastasis
CC or as regulators of cell migration for transplantation therapy, and also
CC for controlling the mobility and cell death of nerve cells, promoting
CC decomposition of the actin-binding protein e.g. filamin-interacting
CC protein in the treatment of preinfiltricular nodular heterotopia
XX

SQ Sequence 965 AA;

Alignment Scores:

	Pred. No.:	0	Length:	965
	Score:	4838.00	Matches:	965
	Percent Similarity:	100.00%	Conservative:	0
	Best Local Similarity:	100.00%	Mismatches:	0
	Query Match:	62.58%	Indels:	0
	DB:	6	Gaps:	0
	US-10-788-793-1 (1-4364) x ABP97030 (1-965)			
QY	816	ATGTTGGTGGACGAGAGGCAGATGCACATCGAGCAACTGGGCCTGCAGAGTCAGAAAGTC	875	
Db	1	MetLeuValAspGluArgGlnMetHisIleGluGlnLeuGlyLeuGlnSerGlnLysVal	20	
QY	876	CAGGACCTCACTCAGAAAGCTGAGGGAGGAGGAAGAAAAAAGCTCAAAAGCGTCACTTACAAA	935	
Db	21	GlnAspLeuThrGlnLysLeuArgGluGluGluGluLysLeuLysAlaValThrTyrLys	40	
QY	936	TCCAAGGAAGACCGCCAGAAAGCTGCTCAAGTTAGAAAGTGGAAGCTTCGAACACAAGGCCTCG	995	
Db	41	SerLysGluAspArgGlnLysLeuLeuLysLeuGluValAspPheGluHisLysAlaSer	60	
QY	996	AGGTTTCCCAGGAGCACGAAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCACAAAC	1055	
Db	61	ArgPheSerGlnGluHisGluGluMetAsnAlaLysLeuAlaAsnGlnGluSerHisAsn	80	
QY	1056	CGGCAACTTCGACTCAAACTGGTTGGCTTATCGCAAGGATTGAGGAGTGGAAAGAGACC	1115	
Db	81	ArgGlnLeuArgLeuLysLeuValGlyLeuSerGlnArgIleGluGluLeuGluGluThr	100	
QY	1116	AATAAAAGCCTTCAGAAGGCAGAGGAAGAGACTCCAGGAGCTGAGAGAGAAAATTGCCAAA	1175	
Db	101	AsnLysSerLeuGlnLysAlaGluGluGluLeuGlnLeuArgGluLysIleAlaLys	120	
QY	1176	GGGGAATGTGAAACTCCAGTCTCATGGCGGAAGTGAGAGAGTCTGCGCAAGCGCGTGCTT	1235	
Db	121	GlyGluCysGlyAsnSerSerLeuMetAlaGluValGluSerLeuArgLysArgValLeu	140	
QY	1236	GAGATGGAGGGCAAGGATGAAGAGATCACGAAGACCGAGGCCCGAGCTGGAAG	1295	
Db	141	GluMetGluGlyLysAspGluGluIleThrLysThrGluAlaGlnCysArgGluLeuLys	160	
QY	1296	AAGAAGCTCCAAGAGGAAGAACACCCACAGCAAGGAAGCTTAGACTAGAAAGTGGAGAAAGCTG	1355	
Db	161	LysLysLeuGlnGluGluGluHisSerLysGluLeuArgLeuGluValGluLysLeu	180	
QY	1356	CAGAAGAGGATGCTCTGAGCTGGAGAAAGCTGGAGAAAGCGTTTCAGCCGGAGTAAGTCGGAA	1415	
Db	181	GlnLysArgMetSerGluLeuGluLysLeuGluGluAlaPheSerArgSerLysSerGlu	200	
QY	1416	TGCACCCAGCTCCATCTGAACCTGGAGAAGGAGAGAACCTTAACCAAAGACCTGCTGTAAC	1475	
Db	201	CysThrGlnLeuHisLeuAsnLeuGluLysGluLysAsnLeuThrLysAspLeuLeuAsn	220	
QY	1476	GAGCTGGAGGTGTCGAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAG	1535	
Db	221	GluLeuGluValValLysSerArgValLysGluLeuGluCysSerGluSerArgLeuGlu	240	
QY	1536	AAGGCCGAGTTAAGCCTCAAAGATGACCTTACAAAAGCTGAAGTCCTTCACTGTGATGCTG	1595	
Db	241	LysAlaGluLeuSerLeuLysAspAspLeuThrLysLeuLysSerPheThrValMetLeu	260	
QY	1596	GTGATGAGAGGAAAAAATATGATGAGAAAAATAAGCAAGAGAGAGGAAAGTGATGGG	1655	
Db	261	ValAspGluArgLysAsnMetMetCgluLysIleLysGlnGluGluArgLysValAspGly	280	
QY	1656	TTGNAATAAAACCTTAAGGTGGAGCAGGGAAAAAGTCATGGATGTACGGGAAAAAGCTAATC	1715	
Db	281	LeuAsnLysAsnPheLysValGluGlnGlyLysValMetAspValThrGluLysLeuIle	300	
QY	1716	GAGGAAAGCAAGAGCTTTTAAAACTCAAATCTGAAATGGAGGAAAAGGAGTACAGTCTG	1775	
Db	301	GluGluSerLysLysLeuLeuLysLeuLysSerGluMetGluGluLysGluTyrSerLeu	320	

QY 1776 ACAAAGGAGGGATGAGCTGATGGGTAAACTGAGGACGGAAGAAAGGTCCTGTGAA 1835
Db 321 ThrLysGluArgAspGluLeuMetGlyLysLeuArgSerGluGluArgSerCysGlu 340
QY 1836 CTGAGCTGCAGTGTAGACTTACTAAAGAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGG 1895
Db 341 LeuSerCysSerValAspLeuLeuLysLysArgLeuAspGlyIleGluGluValGluArg 360
QY 1896 GAAATAAACCGAGGTAGTCGTGCAAGGGGTCTGAGTTCACCTGCCCGGAAGACAATAAG 1955
Db 361 GluIleAsnArgGlyArgSerCysLysGlySerGluPheThrCysProGluAspAsnLys 380
QY 1956 ATCAGAGAACTAACGCTTGAATTCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTG 2015
Db 381 IleArgGluLeuThrLeuGluIleGluArgLeuLysLysArgLeuGlnGlnLeuGluVal 400
QY 2016 GTGGAGGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAGCAGAAGTTCAGA 2075
Db 401 ValGluGlyAspLeuMetLysThrGluAspGluTyrAspGlnLeuGluGlnLysPheArg 420
QY 2076 ACCGAGCAGGATAAGGCAAAACCTCCTCTCCCAGCAGCTCGAGGAAATCAACACACCAAATG 2135
Db 421 ThrGluGlnAspLysAlaAsnPheLeuSerGlnGlnLeuGluGluIleLysHisGlnMet 440
QY 2136 GCCAAGCACAAAGCCATAGAGAAAGGGAGGCCGTGAGCCAGGAAGCCGAACCTGCGACAC 2195
Db 441 AlaLysHisLysAlaIleGluLysGlyGluAlaValSerGlnGluAlaGluLeuArgHis 460
QY 2196 AGGTTTCGGCTGGAGAGGCTAAAGTTCGTGATTACAGGCCGAGGTGCAGGCTCTCAAG 2255
Db 461 ArgPheArgLeuGluAlaLysSerArgAspLeuGlnAlaGluValGlnAlaLeuLys 480
QY 2256 GAGAAGATCCACGAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTAT 2315
Db 481 GluLysIleHisGluLeuMetAsnLysGluAspGlnLeuSerGlnLeuGlnValAspTyr 500
QY 2316 TCGGTCCTTCAGCAAAAGATTATGGAAGAAGAAACTAAGAACAAAGACATGGGAGGGAG 2375
Db 501 SerValLeuGlnGlnArgPheMetGluGluThrLysAsnLysAsnMetGlyArgGlu 520
QY 2376 GTCCTCAATCTGACCAAGGAGCTAGAGCTTCCAGCGCTACAGCCGAGCTCTCAGGCCG 2435
Db 521 ValLeuAsnLeuThrLysGluLeuGluLeuSerLysArgTyrSerArgAlaLeuArgPro 540
QY 2436 AGTGGGAACGGCCGAAGGATGGTGGACGTGCCTGTGGCCTCCACTGGGGTGACACCGAG 2495
Db 541 SerGlyAsnGlyArgArgMetValAspValProValAlaSerThrGlyValGlnThrGlu 560
QY 2496 GCGGTGTGCGGGGATGCTCGGAGGAGGAGACCCCGGCTGTGTTTCATTTCGCAAAATCCTTC 2555
Db 561 AlaValCysGlyAspAlaAlaGluGluThrProAlaValPheIleArgLysSerPhe 580
QY 2556 CAGGAGGAAAATCATCATGAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCCATGGAA 2615
Db 581 GlnGluGluAsnHisIleMetSerAsnLeuArgGlnValGlyLeuLysLysProMetGlu 600
QY 2616 CGGTCTCGGTCTCGACAGGTATCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCT 2675
Db 601 ArgSerSerValLeuAspArgTyrProProAlaAlaAsnGluLeuThrMetArgLysSer 620
QY 2676 TGGATTCTTGATGAGAAAAAGAGAAAACGGTCTCTCCACTCCGACGAGAAAAGGGCCC 2735
Db 621 TrpIleProTrpMetArgLysArgGluAsnGlyProSerThrProGlnGluLysGlyPro 640
QY 2736 AGGCCAAACCCAGGTCGAGGGCACCCCGGGGAGCTGGTCTCTAGCACCAAGCAGGGCCAG 2795
Db 641 ArgProAsnGlnGlyAlaGlyHisProGlyGluLeuValLeuAlaProLysGlnGlyGln 660
QY 2796 CCCCTACACATCCGTCGTACACACCAGATCATGAGAACAGCAGCTGCCACCCCTGAGATCACA 2855
Db 661 ProLeuHisIleArgValThrProAspHisGluAsnSerThrAlaThrLeuGluIleThr 680
QY 2856 AGCCCCACATCTGAAGAGTTTTTCTCTAGTACCACCGTCATTCTCTACCTTAGGCAACCCAG 2915

Db 681 SerProThrSerGluGluPhePheSerSerThrThrValIleProThrLeuGlyAsnGln 700
QY 2916 AAACCAAGAATAACCATATTATTCATCACCCCAATGTCTGCGAAAAGCCCAAAAGTGCA 2975
Db 701 LysProArgIleThrIleIleProSerProAsnValMetSerGlnLysProLysSerAla 720
QY 2976 GATCCTACTCTCGGCCCCAGAAACGAGCCCATGTCCCTGTCTACGATTACTACTATTCCAGA 3035
Db 721 AspProThrLeuGlyProGluArgAlaMetSerProValThrIleThrThrIleSerArg 740
QY 3036 GAGAAGAGCCCGGAAGGTGGAAGGAGCGCCTTTGCCGACAGGCCTGCATCCCCCATCCAA 3095
Db 741 GluLysSerProGluGlyArgSerAlaPheAlaAspArgProAlaSerProIleGln 760
QY 3096 ATCATGACGGTGTCAACATCTGCAGCTCCCAGTCCCAGTGAATCGTGTCTCTCTGAATCTCAG 3155
Db 761 IleMetThrValSerThrSerAlaAlaProThrGluIleAlaValSerProGluSerGln 780
QY 3156 GAAGTGCCTATGGGAAGGACTATCCTCAAAGTCACCCCGGAAAAACAACTGTTCAGGCC 3215
Db 781 GluValProMetGlyArgThrIleLeuLysValThrProGluLysGlnThrValProAla 800
QY 3216 CCCGTGCGGAAGTACAACTCCAATGCTAATATCATCATCACCGGAAGACATAAAATTAC 3275
Db 801 ProValArgLysTyrAsnSerAsnAlaAsnIleIleThrThrGluAspAsnLysIleHis 820
QY 3276 ATTCACCTGGGTCTCAGTTTAAGCGATCTCCTGGGCGCTGCCGCTGAAGGCGTGAGCCCCA 3335
Db 821 IleHisLeuGlySerGlnPheLysArgSerProGlyProAlaAlaGluGlyValSerPro 840
QY 3336 GTTATCACCGTCCGGCCTGTCAACGTGACAGCGGAGAGAGGTTTCTACAGGCACAGTC 3395
Db 841 ValIleThrValArgProValAsnValThrAlaGluLysGluValSerThrGlyThrVal 860
QY 3396 CTTGCTCTCCAGGAACCACTCTCTTCAAGACCCCGGTGCTAGCAAAAGTGACACAGCACT 3455
Db 861 LeuArgSerProArgAsnHisLeuSerSerArgProGlyAlaSerLysValThrSerThr 880
QY 3456 ATAACTATAACCCCGGTCAACAACGTTCATCCACACGAGGAACCCCAATCAGTGTGAGGACAA 3515
Db 881 IleThrIleThrProValThrThrSerSerThrArgGlyThrGlnSerValSerGlyGln 900
QY 3516 GATGGGTCTATCTCAGCGGCCTACCCCCACCCGCATTCTTATGTCAAAGGTATGAAAGCT 3575
Db 901 AspGlySerSerGlnArgProThrProThrArgIleProMetSerLysGlyMetLysAla 920
QY 3576 GGAAGCCAGTAGTGGCAGCCTCAGGAGCAGGAAATCTGACCAAATTCAGCCTCGAGCT 3635
Db 921 GlyLysProValValAlaAlaSerGlyAlaGlyAsnLeuThrLysPheGlnProArgAla 940
QY 3636 GAGACTCAGTCTATGAAAAATAGAGCTGAAGAAAATCTGCAGCCAGCAGCTGCCTCTCTT 3695
Db 941 GluThrGlnSerMetLysIleGluLeuLysLysSerAlaAlaSerSerThrAlaSerLeu 960
QY 3696 GGAGGGGGGAAGGC 3710
Db 961 GlyGlyGlyLysGly 965
RESULT 5
AAB93085
ID AAB93085 standard; protein; 653 AA.
XX
AC AAB93085;
XX
XX 26-JUN-2001 (first entry)
DT
XX Human protein sequence SEQ ID NO:11924.
DE
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
KW
XX Homo sapiens.
OS
XX

PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX PF 29-JUL-1999; 99JP-00248036.
XX PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
PA
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
DR
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
PT
XX
PS Claim 8; SEQ ID NO 11924; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
XX
SQ Sequence 653 AA;

Alignment Scores:
Pred. No.: 1.68e-198 Length: 653
Score: 3051.00 Matches: 603
Percent Similarity: 95.25% Conservative: 19
Best Local Similarity: 92.34% Mismatches: 31
Query Match: 39.46% Indels: 0
DB: 4 Gaps: 0

US-10-788-793-1 (1-4364) x AAB933085 (1-653)

QY 1752 ATGGAGGAAAAGGAGTACAGTCTGACAAAGGAGGAGGATGAGTGATGGTTAAACTGAGG 1811
Db 1 MetGluGluLysValTyrAsnLeuThrArgGluArgAspGluLeuIleGlyLysLeuLys 20

QY 1812 AGCGAAGAAGAGGTCCTGTGAACTGAGCTGAGTGTAGACTTACTAAAGAAGCGGCTT 1871
Db 21 SerGluGluGluLysSerSerGluLeuSerCysSerValAspLeuLeuLysLysArgLeu 40

QY 1872 GATGGCATAGAGGAGGTAGAAAGGGAATAAACCAGGAGGTAGGTCGTGCAAGGGCTCTGAG 1931
Db 41 AspGlyIleGluGluValGluArgGluIleThrArgGlyArgSerArgLysGlySerGlu 60

QY 1932 TTCACCTGCCCGAAGACAATAAGATCAGAGAACTAACGCTTGAAATCGAGAGACTGAAG 1991
Db 61 LeuThrCysProGluAspAsnLysIleLysGluLeuThrLeuGluIleGluArgLeuLys 80

QY 1992 AAACGGCTCCAGCAGTTGAGGTTGGTGGAGGGGACTTGATGAAGACCCGAGACGAATAT 2051
Db 81 LysArgLeuGlnGlnLeuGluValValGluGlyAspLeuMetLysThrGluAspGluTyr 100

QY 2052 GACCAGTTGGAGCAGAAAGTTCAGAACCCGACGAGGATAAGSCAAACTTCCTCTCCAGCAG 2111
Db 101 AspGlnLeuGluGlnLysPheArgThrGluGlnAspLysAlaAsnPheLeuSerGlnGln 120

QY 2112 CTCGAGGAAATCAAACACCAAATGGCCAAAGCACAAGCCATAGAGAAAGGGAGGCCGTG 2171
Db 121 LeuGluGluIleLysHisGlnIleAlaLysAsnLysAlaIleGluLysGlyGluValVal 140

QY 2172 AGCCAGGAAGCCGAACCTCGACACAGGTTTCGGCTGGAGGAGGCTAAAAGTCGTGATTTA 2231
Db 141 SerGlnGluAlaGluLeuArgHisArgPheArgLeuGluGluAlaLysSerArgAspLeu 160

QY 2232 CAGGCCGAGGTGCAGGCTCTCAAGGAGAGAGATCCACGAGCTGATGAACAAGGAAGACCAG 2291
Db 161 LysAlaGluValGlnAlaLeuLysGluLysIleHisGluLeuMetAsnLysGluAspGln 180

QY 2292 CTGTCTCAGCTCCAAGTCGACTATTTCGGTCTTCAGCAAAAGATTTATGGAAGAAACT 2351
Db 181 LeuSerGlnLeuGlnValAspTyrSerValLeuGlnGlnArgPheMetGluGluGluAsn 200

QY 2352 AAGAACAAAGACATGGGGAGGAGGTCTCTCAATCTGACCAAGGAGCTAGAGCTTCCAAG 2411
Db 201 LysAsnLysAsnMetGlyGlnGluValLeuAsnLeuThrLysGluLeuGluLeuSerLys 220

QY 2412 CGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAAGGATGGTGACGTGCCTGTG 2471
Db 221 ArgTyrSerArgAlaLeuArgProSerValAsnGlyArgArgMetValAspValProVal 240

QY 2472 GCCTCCACTGGGTGCAGACCGAGCGGTGTGCGGGATGTGCGGAGGAGGAGACCCCG 2531
Db 241 ThrSerThrGlyValGlnThrAspAlaValSerGlyGluAlaGluGluGluThrPro 260

QY 2532 GCTGTGTTTCATTTCGCAATCCTTCCAGGAGGAAAAATCACATCATGAGTAATCTTCGACAG 2591
Db 261 AlaValPheIleArgLysSerPheGlnGluGluAsnHisIleMetSerAsnLeuArgGln 280

QY 2592 GTAGGCCTGAAGAAACCCATGGAACGGTCTCGGTCTCTCGACAGGTATCCCCAGCAGCG 2651
Db 281 ValGlyLeuLysLysProValGluArgSerSerValLeuAspArgTyrProProAlaAla 300

QY 2652 AATGAGCTCACCATGAGGAAGTCTTGGATTCTTGGATGAGAAAAAGAGAAAACGGTCCT 2711
Db 301 AsnGluLeuThrMetArgLysSerTyrIleProTrpMetArgLysArgGluAsnGlyPro 320

QY 2712 TCCACTCCGAGGAGAAAGGGCCAGGCCAAACAGGGTGCAGGGCACCCCGGGAGCTG 2771
Db 321 SerIleThrGlnGluLysGlyProArgThrAsnSerSerProGlyHisProGlyGluVal 340

QY 2772 GTCCTAGCACCAAGCAGGGCCAGCCCCCTACACATCCGTGTGACACCAGATCATGAGAAC 2831
Db 341 ValLeuSerProLysGlnGlyGlnProLeuHisIleArgValThrProAspHisGluAsn 360

QY 2832 AGCACTGCCACCTGGAGATCACAAGCCCCCACATCTGAAGAGTCTTCTCTAGTACCACC 2891
Db 361 SerThrAlaThrLeuGluIleThrSerProThrSerProThrSerGluGluPheSerSerThrThr 380

QY 2892 GTCATTCTCTACCTTAGGCAACCAAGAAACCAAGATAACCATATTTCATCACCCAATGTC 2951
Db 381 ValIleProThrLeuGlyAsnGlnLysProArgIleThrIleIleProSerProAsnVal 400

QY 2952 ATGTCGCAAAAGCCCCAAAAGTGCAGATCCTTCTCTCGGCCCAAGAACGAGCCATGTCCCCT 3011
Db 401 MetProGlnLysGlnLysSerGlyAspThrThrLeuGlyProGluArgAlaMetSerPro 420

QY 3012 GTCACGATTACTACTATTTCAGAGAGAAGAGCCCGGAAGGTGGAAGGAGCGCCTTGCC 3071

Db 421 ValThrIleThrThrPheSerArgGluLysThrProGluSerGlyArgGlyAlaPheAla 440
QY 3072 GACAGGCCTGCATCCCCCATCCAAATCATGACGGTGTCAACATCTGCAGCTCCCACTGAA 3131
Db 441 AspArgProThrSerProIleGlnIleMetThrValSerThrSerAlaAlaProAlaGlu 460
QY 3132 ATCGCTGTCTCTCTGAATCTCAGGAAGTGCCTATGGGAAGGACTATCCTCAAAGTCACC 3191
Db 461 IleAlaValSerProGluSerGlnGluMetProMetGlyArgThrIleLeuLysValThr 480
QY 3192 CCGGAAAAACAACCTGTTCCAGCCCCCGTGGGGAAGTACAACCTCCAATGCTAATATCATC 3251
Db 481 ProGluLysGlnThrValProThrProValArgLysTyrAsnSerAsnAlaAsnIle 500
QY 3252 ACCACGGAAGACAATAAATTACATTCACCTGGTCTCAGTTTAAGCATCTCTCTGGG 3311
Db 501 ThrThrGluAspAsnLysIleHisIleHisLeuGlySerGlnPheLysArgSerProGly 520
QY 3312 CCTGCCGTGAAGCGGTGAGCCAGTTCATCACCGTCCGGCCTGTCAACGTGACAGCGGAG 3371
Db 521 ThrSerGlyGluGlyValSerProValIleThrValArgProValAsnValThrAlaGlu 540
QY 3372 AAGGAGGTTTCTACAGGCACAGTCTCTCGCTCTCCAGGAACCACTCTCTTCAAGACCC 3431
Db 541 LysGluValSerThrGlyThrValLeuArgSerProArgAsnHisLeuSerSerArgPro 560
QY 3432 GGTGCTAGCAAAGTGACCACTATACTATAACCCCGGTGCACAACGTCTATCCACACGA 3491
Db 561 GlyAlaSerLysValThrSerThrIleThrIleThrProValThrThrSerSerAlaArg 580
QY 3492 GGAACCCAAATCAGTGTGAGGACAAGATGGTCTATCTCAGCGGCCTACCCACCCGCATT 3551
Db 581 GlyThrGlnSerValSerGlyGlnAspGlySerSerGlnArgProThrProThrArgIle 600
QY 3552 CCTATGTCAAAGGTATGAAAGCTGGAAGCCAGTAGTGGCAGCCTCAGGACGAGGAAT 3611
Db 601 ProMetSerLysGlyMetLysAlaGlyLysProValValAlaAlaProGlyAlaGlyAsn 620
QY 3612 CTGACCAAATTCAGCCTCGAGCTGAGACTCAGTCTATGAAATAGAGCTGAAGAAATCT 3671
Db 621 LeuThrLysPheGluProArgAlaGluThrGlnSerMetLysIleGluLeuLysLysSer 640
QY 3672 GCAGCCAGCAGCACTGCCCTCTCTTGGAGGGGGGAAGGCG 3710
Db 641 AlaAlaSerSerThrThrSerLeuGlyGlyGlyLysGly 653

RESULT 6
AAB93244
ID AAB93244 standard; protein; 551 AA.
XX
AC AAB93244;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12253.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX

PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX
PS Claim 8; SEQ ID NO 12253; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
XX
SQ Sequence 551 AA;

Alignment Scores:
Pred. No.: 6.21e-167 Length: 551
Score: 2586.00 Matches: 521
Percent Similarity: 98.18% Conservative: 19
Best Local Similarity: 94.73% Mismatches: 10
Query Match: 33.45% Indels: 0
DB: 4 Gaps: 0

US-10-788-793-1 (1-4364) x AAB93244 (1-551)
QY 816 ATGTTGGTGGACGAGAGGCGAGATGCACATCGAGCAACTGGGCTGCAGAGTCAGAAAGTC 875
Db 1 MetLeuValAspGluArgGlnMetHisIleGluGlnLeuGlyLeuGlnSerGlnLysVal 20
QY 876 CAGGACCTCACTCAGAAAGCTGAGGGAGGAGGAAGAAACTCAAAGCGGTCACTTACAA 935
Db 21 GlnAspLeuThrGlnLysLeuArgGluGluGluLysLeuLysAlaIleThrSerLys 40
QY 936 TCCAAGGAAGACCGCCAGAGCTGCTCAAGTTAGAAAGTGGACTTCGAACACAAAGGCCTCG 995
Db 41 SerLysGluAspArgGlnLysLeuLeuLysLeuGluValAspPheGluHisLysAlaSer 60
QY 996 AGGTTTTCCAGGAGCAGCAGAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCACAAC 1055
Db 61 ArgPheSerGlnGluHisGluGluMetAsnAlaLysLeuAlaAsnGlnGluSerHisAsn 80
QY 1056 CGGCAACTTCGACTCAAACTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGAAGAGACC 1115
Db 81 ArgGlnLeuArgLeuLysLeuValGlyLeuThrGlnArgIleGluGluLeuGluGluThr 100
QY 1116 AATAAAAGCCTTCAGAAAGGCGAGGAGAGAGCTCCAGGAGCTGAGAGAGAAAATTCGCAAA 1175
Db 101 AsnLysAsnLeuGlnLysAlaGluGluGluLeuGlnGluLeuArgAspLysIleAlaLys 120

QY 1176 GGGGAATGTGGAAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGGAAGCGGTGCTT 1235
Dbb 121 GlyGluCysGlyAsnSerSerLeuMetAlaGluValGluAsnLeuArgLysArgValLeu 140
QY 1236 GAGATGGAGGGCAAGGATGAAGATCAGCAAGACCGAGGCCAGTCGCCGGAGCTGAAG 1295
Dbb 141 GluMetGluGlyLysAspGluGluIleThrLysThrGluSerGlnCysArgGluLeuArg 160
QY 1296 AAGAAGCTCCAAGAGGAAGAACACCACAGCAAGGAACCTTAGACTAGAAAGTGAGAAAGCTG 1355
Dbb 161 LysLysLeuGlnGluGluGluHisHisSerLysGluLeuArgLeuGluValGluLysLeu 180
QY 1356 CAGAAGAGGATGTCTGAGCTGGAGAAAGCTGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAA 1415
Dbb 181 GlnLysArgMetSerGluLeuGluLysLeuGluAlaPheSerLysSerLysSerGlu 200
QY 1416 TGCACCCAGCTCCATCTGAACCTGGAGAGGAGAAAGAACCTTAACCAAGACCTTGCTGAAC 1475
Dbb 201 CysThrGlnLeuHisLeuAsnLeuGluLysGluLysAsnLeuThrLysAspLeuLeuAsn 220
QY 1476 GAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAG 1535
Dbb 221 GluLeuGluValValLysSerArgValLysGluLeuGluCysSerGluSerArgLeuGlu 240
QY 1536 AAGCCGAGTTAAGCTCAAGATGACCTTACAAAGCTGAAGTCTTCACTGTGATGCTG 1595
Dbb 241 LysAlaGluLeuSerLeuLysAspAspLeuThrLysLysLeuLysSerPheThrValMetLeu 260
QY 1596 GTGGATGAGAGGAAAAATATGATGGAGAAAAATAAAGCAAGAGAGAGAAAGTGGATGG 1655
Dbb 261 ValAspGluArgLysAsnMetMetGluLysIleLysGlnGluArgLysValAspGly 280
QY 1656 TTGAATAAAACTTTAAGTGGAGCAGGGAAGATCATGATGTGACGGAAAAAGCTAATC 1715
Dbb 281 LeuAsnLysAsnPheLysValGluGlnGlyLysValMetAspValThrGluLysLeuIle 300
QY 1716 GAGGAAAGCAAGAAGCTTTTAAAACTCAAACTCGAAATGAGGAAAAAGGAGTACAGTCTG 1775
Dbb 301 GluGluSerLysLysLeuLysLeuLysSerGluMetGluLysValTyraAsnLeu 320
QY 1776 ACAAGGAGAGGGATGAGCTGATGGTAACTGAGGAGCGAAGAAAGATCCTGTGAA 1835
Dbb 321 ThrArgGluArgAspGluLeuIleGlyLysLeuLysSerGluGluGluLysSerSerGlu 340
QY 1836 CTGAGCTGCAGTGTAGACTTACTAAAGAAGCGGCTTGATGGCATAGAGGAGTGAAGG 1895
Dbb 341 LeuSerCysSerValAspLeuLeuLysLysArgLeuAspGlyIleGluGluValGluArg 360
QY 1896 GAAATAAACCGAGGTAGGTCGTCGCAAGGGTCTGAGTTACCTGCCCGGAGACAATAAG 1955
Dbb 361 GluIleThrArgGlyArgSerArgLysGlySerGluLeuThrCysProGluAspAsnLys 380
QY 1956 ATCAGAGAACTAACGCTTGAATCGAGAGACTGAAGAAACGGCTCCAGAGTTGGAGGTG 2015
Dbb 381 IleLysGluLeuThrLeuGluIleGluArgLeuLysLysArgLeuGlnGlnLeuGluVal 400
QY 2016 GTGGAGGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAGCAGAGTTCAGA 2075
Dbb 401 ValGluGlyAspLeuMetLysThrGluAspGluTyraAspGlnLeuGluGlnLysPheArg 420
QY 2076 ACCGAGCAGGATAAAGGCAAACTTCTCTCCAGCAGCTCGAGGAAATCAAACACCAATG 2135
Dbb 421 ThrGluGlnAspLysAlaAsnPheLeuSerGlnGlnLeuGluGluIleLysHisGlnIle 440
QY 2136 GCCAAGCACAAAGCCATAGAGAAAGGGGAGCGCTGAGCCAGGAGCCGAACCTCGCACAC 2195
Dbb 441 AlaLysAsnLysAlaIleGluLysGlyGluValValSerGlnGlnAlaGluLeuArgHis 460
QY 2196 AGGTTTCGGCTGGAGGAGGCTTAAAGTTCGTGATTACAGGCGGAGGTGCAGGCTCTCAAG 2255
Dbb 461 ArgPheArgLeuGluGluAlaLysSerArgAspLeuLysAlaGlyValGlnAlaLeuLys 480
QY 2256 GAGAAGATCCACGAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTAT 2315

Db 481 GluLysIleHisGluLeuMetAsnLysGluAspGlnLeuSerGlnLeuGlnValAspTyr 500
QY 2316 TCGGTCCTTCAGCAAAAGATTTATGGAAGAAGAAACTAAGAACAAGACATCGGGAGGGAG 2375
Db 501 SerValLeuGlnGlnArgPheMetGluGluAsnLysAsnLysAsnMetGlyGlnGlu 520
QY 2376 GTCCTCAATCTGACCAAGAGACTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCG 2435
Db 521 ValLeuAsnLeuThrLysGluLeuGluLeuSerLysArgTyrSerArgAlaLeuArgPro 540
QY 2436 AGTGGGAACGGCCGAAGGATGGTGGACGTG 2465
Db 541 SerValAsnGlyLysArgMetValAspVal 550
RESULT 7
ABR62252
ID ABR62252 standard; protein; 1133 AA.
XX
AC ABR62252;
XX 22-SEP-2003 (first entry)
XX GPBP-interacting protein GIP130b.
DE Goodpasture antigen binding protein interacting protein; GPBP; GIP130b;
KW human; transcription factor; autoimmune disease; cancer;
KW immunosuppressive; cytostatic.
XX Homo sapiens.
OS
XX WO2003048193-A2.
PN
XX 12-JUN-2003.
PD
XX 05-DEC-2002; 2002WO-EP013802.
PF
XX 07-DEC-2001; 2001US-0338287P.
PR 20-MAY-2002; 2002US-0382004P.
PR
XX (SAUS/) SAUS J.
PA (REVE/) REVERT-ROS F.
PA
XX Saus J, Revert-Ros F;
PI
XX WPI; 2003-505281/47.
DR N-PSDB; ACC83929.
DR
XX New Goodpasture antigen binding protein-interacting 90 and 130 kDa
PT polypeptides, useful for diagnosing and/or treating disorders associated
PT with the GIP90/130 polypeptide, such as autoimmune disorders and/or
PT cancer.
XX
XX Claim 8; Page 78-83; 115pp; English.
PS
XX The present sequence is the protein sequence of GIP130b, a novel 130 kDa
CC Goodpasture antigen binding protein (GPBP) interacting protein. The
CC DOC/GIP90/130 mRNA family results from a complex diversification
CC mechanism operating on the expression of the GIP90 gene on chromosome 3
CC (3q12). GIP90/130 polypeptides of the invention interact with GPBP and
CC are capable of aggregation. They can be used to modify GPBP-GIP90/130
CC interactions, to modify Gp90/130 aggregation, and to modulate gene
CC expression. The invention provides GIP90/130 polypeptides, portions of
CC them, antibodies, nucleic acid sequences, expression vectors and host
CC cells, as well as methods for detecting GIP90/130 polypeptides and
CC nucleic acids, and methods for treating an autoimmune disease or cancer
CC by modifying the expression or activity of one or more GIP90/130
CC polypeptides
XX
SQ Sequence 1133 AA;
Alignment Scores:
Pred. No.: 9.53e-158 Length: 1133

Score:	2452.50	Matches:	542
Percent Similarity:	64.40%	Conservative:	214
Best Local Similarity:	46.17%	Mismatches:	337
Query Match:	31.72%	Indels:	81
DB:	6	Gaps:	19
US-10-788-793-1 (1-4364) x ABR62252 (1-1133)			
QY	75	ATGAGATCACGAAATCAAGGTGGAGAAAAGTTTCATCTAACGGGCATGTCTCCTGCCCAAG	134
Db	1	MetArgSerArg-----GlySerAspThrGluGlySerAlaGlnLysLysPheProArg	18
QY	135	TCCTCCATCATCAGCAGTGATGGTGAAGGGCCCTCAGAAGATGCAAAAAAGAACAG	194
Db	19	HisThr-----LysGlyHisSerPheGlnGlyProLysAsnMet	31
QY	195	GCCAATCGGAAGGAGGAG---GATGTCATGGCTTCGGGAACCTATCAAAAGGCACCTCAAA	251
Db	32	LysHisArgGlnGlnAspLysAspSerProSerGluSerAspVal-----IleLeu	48
QY	252	CCATCTGGAGAAAGTGAGAAA-----AAGACTAAGAAGTCTGTGGAGTTATCC	299
Db	49	ProCysProLysAlaGluLysProHisSerGlyAsnGlyHisGlnAlaGluAspLeuSer	68
QY	300	AAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGAGTTGCAGGCTCGAGAAGAT	359
Db	69	ArgAspAspLeuLeuPheLeuLeuSerIleLeuGluGlyGluLeuGlnAlaArgAspGlu	88
QY	360	GTCATCCACATGCTGAGGACACAGAAAAACCAAGCCCGAGGTTCTGGAGGCACACTATGGA	419
Db	89	ValIleGlyIleLeuLysAlaGluLysMetAspLeuAlaLeuLeuGluAlaGlnTyrGly	108
QY	420	TCTGCAGAAACCTGAGAAAGTGCTTCGGGTCTCGACCGAGATGCCATCCTTGCTCAAGAG	479
Db	109	PheValThrProLysLysValLeuGluAlaLeuGlnArgAspAlaPheGlnAlaLysSer	128
QY	480	AAGTCCATAGGAGAAGACGTCTATGAGAAAACCTATCTCAGAGCTGGACAGACTGGAGAA	539
Db	129	ThrProTrpGlnGluAspIleTyrGluLysProMetAsnGluLeuAspLysValValGlu	148
QY	540	AAGCAGAAGGAGACGTACCGCGCATGCTAGAGCAGCTGCTGGCTGAGAAAGTGTCAAC	599
Db	149	LysHisLysGluSerTyrArgArgIleLeuGlyGlnLeuLeuValAlaGluLysSerHis	168
QY	600	AGCGCACCGGTGTACGAGCTGAGAACGAGAACGACACACACTGACTACATGAAACAAG	659
Db	169	ArgGlnThrIleLeuGluLeuGluGluGluLysArgLysHisLysGluTyrMetGluLys	188
QY	660	ACGACGACATTCAACCAACCTGCTGGAGCAGGAGCGAGAGAGGTTGAAAAAGCTCCTTGAA	719
Db	189	SerAspGluPheIleCysLeuLeuGluGlnGluCysGluArgLeuLysLysLeuIleAsp	208
QY	720	CAAGAAAAAGCTTACCAAGCCCGCAAGAAAAAGGAAAAACGCTAAGCGGCTCAACAAACTT	779
Db	209	GlnGluIleLysSerGlnGluGluLysGluGlnGluLysGluLysArgValThrThrLeu	228
QY	780	CGAGATGAGCTTGTGAAGCTCAAGTCTCTTCGCCCTCATGTTGGTGGACGAGAGCAGATG	839
Db	229	LysGluGluLeuThrLysLeuLysSerPheAlaLeuMetValValAspGluGlnArg	248
QY	840	CACATCGAGCAACTGGGCTTCAGAGTCCAGAAAGTCCAGGACCTCACTCAGAAGCTGAGG	899
Db	249	LeuThrAlaGlnLeuThrLeuGlnArgGlnLysIleGlnGluLeuThrThrAsnAlaLys	268
QY	900	GAGGAGGAAGAAAACTCAAGCGGCTCACTTACAATCCAAAGGAAGACCGCCAGAGCTG	959
Db	269	GluThrHisThrLysLeuAlaLeuAlaGluAlaArgValGlnGluGluGlnLysAla	288
QY	960	CTCAAGTTAGAAGTGGACTTCGAAACACAAGGCCTCGAGGTTTCCAGGAGCAGCAAGAG	1019
Db	289	ThrArgLeuGluLysGluLeuGlnThrGlnThrThrLysPheHisGlnAspGlnAspThr	308
QY	1020	ATGAACGCCAAATTGGCGAATCAAGAATCTCACACCGGCAACTTCGACTCAAACTGGTT	1079

Db	309	IleMetAlaLysLeuThrAsnGluAspSerGlnAsnArgGlnLeuGlnLysLeuAla	328
QY	1080	GGCTTATCGCAAAGGATTGAGGAGCTGGAAGAGACCAATAAAGCCCTTCAGAAGGCAGAG	1139
Db	329	AlaLeuSerArgGlnIleAspGluLeuGluThrAsnArgSerLeuArgLysAlaGlu	348
QY	1140	GAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAACCTCCAGTCTC	1199
Db	349	GluGluLeuGlnAspIleLysGluLysIleSerLysGlyGluTyrGlyAsnAlaGlyIle	368
QY	1200	ATGGCGGAAGTGGAGAGTCTCGCAAGCGCGTCTTGAGATGGAGGGCAAGGATGAAGAG	1259
Db	369	MetAlaGluValGluLeuArgLysArgValLeuAspMetGluGlyLysAspGluGlu	388
QY	1260	ATCACGAAGACCGAGGCCAGTCCGGGAGCTGGAAGAAAGCTCCAAGAGGAAGAACAC	1319
Db	389	LeuIleLysMetGluGlnCysArgAspLeuAsnLysArgLeuGluArgGluThrLeu	408
QY	1320	CACAGCAAGGAACCTAGACTAGAAGTGGAGAGCTGCAGAAGAGGATGTCTGAGCTGGAG	1379
Db	409	GlnSerLysAspPheLysLeuGluValGluLysLeuSerLysArgIleMetAlaLeuGlu	428
QY	1380	AAGCTGGAGGAAGCTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTG	1439
Db	429	LysLeuGluAspAlaPheAsnLysSerLysGlnGluCysTyrSerLeuLysCysAsnLeu	448
QY	1440	GAGAAGGAGAAGAACCTAACCAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGA	1499
Db	449	GluLysGluArgMetThrThrLysGlnLeuSerGlnGluLeuGluSerLeuLysValArg	468
QY	1500	GTAAAGAACTCGAATGCTCCGAGAGTAGACTGAGAGGCGCCGAGTTAAGCCCTCAAGAT	1559
Db	469	IleLysGluLeuGluAlaIleGluSerArgLeuGluLysThrGluPheThrLeuLysGlu	488
QY	1560	GACCTTACAAAGCTGAAAGTCTTCACTGTGATGCTGCTGGATGAGAGGAAAAATATGATG	1619
Db	489	AspLeuThrLysLeuLysThrLeuThrValMetPheValAspGluArgLysThrMetSer	508
QY	1620	GAGAAAAATAAGCAAGAGAGAGGAAAGTGGATGGTTGAATAAAAACTTTAAGGTGGAG	1679
Db	509	GluLysLeuLysLysThrGluAspLysLeuGlnAlaAlaSerSerGlnLeuGlnValGlu	528
QY	1680	CAGGAAAAAGTCATGGATGTGACGGAAGAAAGCTAATCGAGGAAAGCAAGAGCTTTAAAA	1739
Db	529	GlnAsnLysValThrThrValThrGluLysLeuIleGluGluThrLysArgAlaLeuLys	548
QY	1740	CTCAAACTCTGAAATGGAGGAAAGAGTACACTGTACAAAGGAGAGGGATGAGCTGATG	1799
Db	549	SerLysThrAspValGluGluLysMetTyrSerValThrLysGluArgAspAspLeuLys	568
QY	1800	GGTAAACTGAGGACGGAAGAAAGGTCCTGTGAACTGAGCTGCAGTGTAGACTTACTA	1859
Db	569	AsnLysLeuLysAlaGluGluLysGlyAsnAspLeuLeuSerArgValAsnMetLeu	588
QY	1860	AAGAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGGGAAATAAACCGAGGTAGG-----	1913
Db	589	LysAsnArgLeuGlnSerLeuGluAlaIleGluLysAspPheLeuLysAsnLysLeuAsn	608
QY	1914	-----TCGTGCAAGGGGTCTGAGTTACCTTCACCTGCCCGGAAGACAATAAGATCAGAGAACTA	1967
Db	609	GlnAspSerGlyLysSerThrThrAlaLeuHisGlnGluAsnAsnLysIleLysGluLeu	628
QY	1968	ACGCTTGAAATCGAGAGACTGAAGAAACCGGCTCCAGCAGTTGGAGGTGGTGGAGGGGAC	2027
Db	629	SerGlnGluValGluArgLeuLysLeuLysLeuLysAspMetLysAlaIleGluAspAsp	648
QY	2028	TTGATGAAGACCGAGGACGAATATGACCAGTGGAGCAGAAAGTTTCAGAACCCAGCAGAT	2087
Db	649	LeuMetLysThrGluAspGluTyrGluThrLeuGluArgArgTyrAlaAsnGluArgAsp	668
QY	2088	AAGGCAAACTTCTCTCCCAGCAGCTCGAGGAATCAAAACCAAAATGGCCAAGCACAAA	2147

Db 669 LysAlaGlnPheLeuSerLysGluLeuGluHisValLysMetGluLeuAlaLysTyrLys 688
Qy 2148 GCCATAGAGAAAGGGGCGGTGAGCCAGGAAGCCGAATCGCGACACAGGTTTCGGCTG 2207
Db 689 LeuAlaGluLysThrGlu---ThrSerHisGluGlnTrpLeuPheLysArgLeuGlnGlu 707
Qy 2208 GAGGAGCTAAAAGTCGTGATTTACAGGCGGAGGTGCAGCTCTCAAGGAGAAGATCCAC 2267
Db 708 GluGluAlaLysSerGlyHisLeuSerArgGluValAspAlaLeuLysGluLysIleHis 727
Qy 2268 GAGCTGATGAACAAGGAGACACGAGCTGTCTCAGCTCCAAGTCGACTATTCGGTCTTCAG 2327
Db 728 GluTyrMetAlaThrGluAspLeuIleCysHisLeuGlnGlyAspHisSerValLeuGln 747
Qy 2328 CAAAGATTATGGAAGAACTAAGAACAAAGAACATGGGAGGGAGGTCTCAATCTG 2387
Db 748 LysLysLeuAsnGlnGlnGluAsnArgAsnArgAspLeuGlyArgGluIleGluAsnLeu 767
Qy 2388 ACCAAGGAGCTAGAGCTTTCCAAGCGCTACAGCCGAGCTTCAGGCCGAGTGGGAACGGC 2447
Db 768 ThrLysGluLeuGluArgTyrArgHisPheSerLysSerLeuArgProSerLeuAsnGly 787
Qy 2448 CGAAGGATGGTGACGTCGCTGTGGCTCCACTGGGGTGCGACCCGAGGGCGTGTGCGGG 2507
Db 788 ArgArgIleSerAspProGlnValPheSerLysGluValGlnThrGluAlaVal----- 805
Qy 2508 GATGCTGCGGAGGAGGAGACCCCG-----GCTGTG 2537
Db 806 -----AspAsnGluProProAspTyrLysSerLeuIleProLeuGluArgAlaVal 822
Qy 2538 TTCATTGCAAAATCCTTCCAGGAG---GAAATACATCATGAGTAATCTTCGACAGGTA 2594
Db 823 IleAsnGlyGlnLeuTyrGluGluSerGluAsnGlnAsp----- 835
Qy 2595 GGCCTGAAGAAACCCATGGAACGGTCTCTCGTCTCTC-----GACAGGTAT 2639
Db 836 -----GluAspProAsnAspGluGlySerValLeuSerPheLysCysSerGlnSerThr 853
Qy 2640 CCCCCAGCAGCAATGAGTCAACCATGAGGAAGTCTTGGATTCCTTGGATGAGAAAAAGA 2699
Db 854 ProCysProValAsn-----ArgLysLeuTrpIleProTrpMetLysSerLys 869
Qy 2700 GAA-----AACGGTCTCTCCACTCCGACGAGGAAGGGCCCCAGGCCAAACCCAG 2747
Db 870 GluGlyHisLeuGlnAsnGly-----LysMetGlnThrLysProAsnAla 884
Qy 2748 GGTGCAGGCGACCCGGGAGTGGTCTTAGCACCAAAGCAGGCGCCAGCCCTACACATC 2807
Db 885 AsnPheValGlnProGlyAspLeuValLeuSerHisThrProGlyGlnProLeuHisIle 904
Qy 2808 CGTGTGACACGATCATGAGAACAGCAGCTGCCACCCCTGGAGATCACAGCCCCACATCT 2867
Db 905 LysValThrProAspHisValGlnAsnThrAlaThrLeuGluIleThrSerProThrThr 924
Qy 2868 GAA-----GAGTTTCTCTAGTACCACCGTCATTCCTACCTTAGGCAACCCAGAAACCA 2921
Db 925 GluSerProHisSerTyrThrSerThrAlaValIleProAsnCysGlyThrProLysGln 944
Qy 2922 AGAATAACCATATTCCATACCCCAATGTCAATGTTCGAAAAGCCCCAAA---AGTGCAGAT 2978
Db 945 ArgIleThrIleLeuGlnAsnAlaSerIleThrProValLysSerLysThrSerThrGlu 964
Qy 2979 CCTACTCTCGGCCCCAGAACGACCCATGTCCCCGTGTCACGATTACTACTATTTCAGAGAG 3038
Db 965 AspLeuMetAsnLeuGluGlnGlyMetSerProIleThrMetAlaThrPheAlaArgAla 984
Qy 3039 AAGAGCCCGGAAGGTGGAAGGACGCGCTTTGGCGACAGGCGCTGCATCCCCCATCCAAATC 3098
Db 985 GlnThrProGluSerCysGlySerLeuThrProGluArgThrMetSerProIleGlnVal 1004
Qy 3099 ATGACGGTGTCAACATCTGACGTCCCACCTGAAATCGTGTCTCTCTGAAATCTCAGGAA 3158
Db 1005 LeuAlaValThrGlySerAlaSerSerProGluGlnGlyArgSerProGluProThrGlu 1024

Qy 3159 GTGCCTATGGGAAGGACTATCTCTCAAAGTCACCCCGGAAAAACAAACTGTTCCAGCCCC 3218
Db 1025 IleSerAlaLysHisAlaIlePheArgValSerProAspArgGlnSerSerTrpGlnPhe 1044
Qy 3219 GTGCGGAAGTACAACCTCCAATGCTAATATCATCACCCACGGAAGACAATAAAATTACATT 3278
Db 1045 GlnArgSerAsnSerAsnSerSerValIleThrThrGluAspAsnLysIleHisIle 1064
Qy 3279 CACCTGGTCTCTCAGTTTAAGCGATCTCTCGGGCCTGCCGTGAAGGCGTGAGCCCCAGTT 3338
Db 1065 HisLeuGlySerProTyrMetGlnAla-----ValAlaSerProValArgProAla 1081
Qy 3339 ATCACCGTCCGGCCTGTCAACGTGACAGCGGAGAAGGAGTTTCTACAGGCACAGTCCCTT 3398
Db 1082 SerProSerAlaProLeuGlnAspAsnArgThrGlnGlyLeuIleAsnGlyAlaLeuAsn 1101
Qy 3399 CGCTCTCCAGGAACCCACTCTCTTCAAGACCCCGGTGCTAGCAAAAGTGACCAAGCACTATA 3458
Db 1102 LysThr-----ThrAsnLysValThrSerSerIle 1111
Qy 3459 ACTATAACCCCGGTCACAACGTCATCCACAGGAGGAACCCAA 3500
Db 1112 ThrIleThrProThrAlaThrProLeuProArgGlnSerGln 1125
RESULT 8
ABR62253
ID ABR62253 standard; protein; 1133 AA.
XX
AC ABR62253;
XX
DT 22-SEP-2003 (first entry)
DE
DE GPBP-interacting protein GIP130c.
XX
KW Goodpasture antigen binding protein interacting protein; GPBP; GIP130c;
KW human; transcription factor; autoimmune disease; cancer;
KW immunosuppressive; cytostatic.
XX
OS Homo sapiens.
PN WO2003048193-A2.
XX
PD 12-JUN-2003.
XX
PF 05-DEC-2002; 2002WO-EP013802.
XX
PR 07-DEC-2001; 2001US-0338287P.
PR 20-MAY-2002; 2002US-0382004P.
XX
PA (SAUS/) SAUS J.
PA (REVE/) REVERT-ROS F.
XX
PI Saus J, Revert-Ros F;
XX
DR WPI; 2003-505281/47.
DR N-PSDB; ACC83930.
XX
PT New Goodpasture antigen binding protein-interacting 90 and 130 kDa
PT polypeptides, useful for diagnosing and/or treating disorders associated
PT with the GIP90/130 polypeptide, such as autoimmune disorders and/or
PT cancer.
XX
PS Claim 9; Page 89-94; 115pp; English.
XX
CC The present sequence is the protein sequence of GIP130c, a novel 130 kDa
CC Goodpasture antigen binding protein (GPBP) interacting protein. The
CC DOC/GIP90/130 mRNA family results from a complex diversification
CC mechanism operating on the expression of the GIP90 gene on chromosome 3
CC (3q12). GIP90/130 polypeptides of the invention interact with GPBP and
CC are capable of aggregation. They can be used to modify GPBP-GIP90/130
CC interactions, to modify GP90/130 aggregation, and to modulate gene
CC expression. The invention provides GIP90/130 polypeptides, portions of

CC them, antibodies, nucleic acid sequences, expression vectors and host
CC cells, as well as methods for detecting GIP90/130 polypeptides and
CC nucleic acids, and methods for treating an autoimmune disease or cancer
CC by modifying the expression or activity of one or more GIP90/130
CC polypeptides
XX
SQ Sequence 1133 AA;

Alignment Scores:

Pred. No.: 3.33e-157 Length: 1133
Score: 2444.50 Matches: 541
Percent Similarity: 64.31% Conservatives: 214
Best Local Similarity: 46.08% Mismatches: 338
Query Match: 31.62% Indels: 81
DB: 6 Gaps: 19

US-10-788-793-1 (1-4364) x ABR62253 (1-1133)

QY	75	ATGAGATCAGCAATCAAGGTGGAGAAAGTTTCATCTAACGGGCATGTCTCTGCCCCCAAG	134
Db	1	MetArgSerArg-----GlySerAspThrGluGlySerAlaGlnLysLysPheProArg	18
QY	135	TCCTCCATCATCAGCAGTGATGGTGAAGGGCCCTCAGAAGATGCAGAAAGAACAAAG	194
Db	19	HisThr-----LysGlyHisSerPheGlnGlyProLysAsnMet	31
QY	195	GCCAAATCGGAAGGAGGAG--GATGTCTATGGCTTCCGGAACTATCAAAAGGCACCTCAAA	251
Db	32	LysHisArgGlnGlnAspLysAspSerProSerGluSerAspVal-----IleLeu	48
QY	252	CCATCTGGAGAAAGTGAGAAA-----AAGACTAAGAAGTCTGTGGAGTTATCC	299
Db	49	ProCysProLysAlaGluLysProHisSerGlyAsnGlyHisGlnAlaGluAspLeuSer	68
QY	300	AAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGGAGTTGCAGGCTCGAGAAGAT	359
Db	69	ArgAspAspLeuLeuPheLeuLeuSerIleLeuGluGlyGluLeuGlnAlaArgAspGlu	88
QY	360	GTCATCCACATGCTGAGACAGAGAGAAAACCAAGCCCGAGGTTCTGGAGGCACACTATGGA	419
Db	89	ValIleGlyIleLeuLysAlaGluLysMetAspLeuAlaLeuLeuGluAlaGlnTyrGly	108
QY	420	TCTGCAGAACCTGAGAAAGTGCTTCGGGTCTGCACCGGAGATGCCATCCTTGCTCAAGAG	479
Db	109	PheValThrProLysLysValLeuGluAlaLeuGlnArgAspAlaPheGlnAlaLysSer	128
QY	480	AAGTCCATAGGAGAGACGCTCTATGAGAAAACCTATCTCAGAGCTGGACAGACTGGAGAA	539
Db	129	ThrProTrpGlnGluAspIleTyrGluLysProMetAsnGluLeuAspLysValValGlu	148
QY	540	AAGCAGAAGGAGACGTACCGCCGCATGCTAGAGAGCTGCTGCTGGCTGAGAAGTGTAC	599
Db	149	LysHisLysGluSerTyrArgArgIleLeuGlyGlnLeuLeuValAlaGluLysSerArg	168
QY	600	AGGCGCACCGTGACAGCTGGAGAACGAGAACGACAGCACACTGACTACATGAACAAG	659
Db	169	ArgGlnThrIleLeuGluLeuGluGluGlnGluCysGluArgLeuLysLysLeuIleAsp	188
QY	660	AGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGAGAGAGGTTGAAAAAGCTCCTGAA	719
Db	189	SerAspGluPheIleCysLeuLeuGluGlnGluCysGluArgLeuLysLysLeuIleAsp	208
QY	720	CAAGAAAAAGCTTACCAGCCCGCAAGAAAAGGAAAAACGCTAAGCGGCTCAACAACTT	779
Db	209	GlnGluIleLysSerGlnGluGluLysGluGlnGluLysGluLysArgValThrThrLeu	228
QY	780	CGAGATGAGCTTGTGAAGCTCAAGTCCCTTCGCCCTCATGTTGGTGGACGAGGCGAGATG	839
Db	229	LysGluGluLeuThrLysLeuLysSerPheAlaLeuMetValValAspGluGlnArg	248
QY	840	CACATCGAGCAACTGGCGCTGCAGAGTCAAGAAAGTCCAGGACCTCACTCAGAAGCTGAGG	899
Db	249	LeuThrAlaGlnLeuThrLeuGlnArgGlnLysIleGlnGluLeuThrThrAsnAlaLys	268

QY	900	GAGGAGGAAGAAAAAAGTCAAAGCGGTCACTTACAAAATCCAAGGAAGACCGCCAGAGCTG	959
Db	269	GluThrHisThrLysLeuAlaLeuAlaGluAlaArgValGlnGluGluGlnLysAla	288
QY	960	CTCAAGTTAGAGTGGACTTCGAACACACAGGCCTCGAGGTTTCCAGGAGCAGCAAGAG	1019
Db	289	ThrArgLeuGluLysGluLeuGlnThrGlnThrThrLysPheHisGlnAspGlnAspThr	308
QY	1020	ATGAACGCCAAATTTGGCGAATCAAGAATCTCACAAACCCGGCAACTTCGACTCAAACTGGTT	1079
Db	309	IleMetAlaLysLeuThrAsnGluAspSerGlnAsnArgGlnLeuGlnGlnLysLeuAla	328
QY	1080	GGCTTATCGAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAGGCAGAG	1139
Db	329	AlaLeuSerArgGlnIleAspGluLeuGluThrAsnArgSerLeuArgLysAlaGlu	348
QY	1140	GAAGAGCTCCAGGAGCTGAGAGAGAGAAAATTTGCCAAAAGGGGAATGTGGAATCTCCAGTCTC	1199
Db	349	GluGluLeuGlnAspIleLysGluLysIleSerLysGlyGluTyrGlyAsnAlaGlyIle	368
QY	1200	ATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTCTTGAGATGGAGGGCAAGGATGAAGAG	1259
Db	369	MetAlaGluValGluLeuArgLysArgValLeuAspMetGluGlyLysAspGluGlu	388
QY	1260	ATCAGGAAGACCCAGGCCAGTGCCTGCGGAGCTGAAGAAGAAGCTCCAAGAGGAACAC	1319
Db	389	LeuIleLysMetGluGlnCysArgAspLeuAsnLysArgLeuGluArgGluThrLeu	408
QY	1320	CACAGCAAGGAACCTTAGACTAGAAGTGGAGAAGCTGCAGAAGAGGATGTCGAGCTGGAG	1379
Db	409	GlnSerLysAspPheLysLeuGluValGluLysLeuSerLysArgIleMetAlaLeuGlu	428
QY	1380	AAGCTGGAGGAAGGCTTCAGCCGGAGTAAAGTCGGGAATGCACCCAGCTCCATCTGAACCTG	1439
Db	429	LysLeuGluAspAlaPheAsnLysSerLysGlnGluCysTyrSerLeuLysCysAsnLeu	448
QY	1440	GAGAAGGAGAGAACCTAACCAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGA	1499
Db	449	GluLysGluArgMetThrThrLysGlnLeuSerGlnGluLeuGluSerLeuLysValArg	468
QY	1500	GTTAAAGAAGTCCGAATGCTCCGAGAGTAGACTGAGAGAGGCCGAGTTAAGCCTCAAAGAT	1559
Db	469	IleLysGluLeuGluAlaIleGluSerArgLeuGluLysThrGluPheThrLeuLysGlu	488
QY	1560	GACCTTACAAAGCTGAAGTCTTCACTGTGATGCTGGTGGATGAGAGGAAAATATATGATG	1619
Db	489	AspLeuThrLysLeuLysThrLeuThrValMetPheValAspGluArgLysThrMetSer	508
QY	1620	GAGAAAATAAGCAAGAGAGAGGAAAGTGGATGGTGTGAATAAAAACTTTAAGGTGGAG	1679
Db	509	GluLysLeuLysLysThrGluAspLysLeuGlnAlaAlaSerSerGlnLeuGlnValGlu	528
QY	1680	CAGGGAAGATCATGGATGTGACGGAAAAGCTAAATCGAGGAAGCAAGAAGCTTTTAAAA	1739
Db	529	GlnAsnLysValThrThrValThrGluLysLeuIleGluGluThrLysArgAlaLeuLys	548
QY	1740	CTCAAATCTGAAATGGAGGAAAAGGAGTACAGTCTGACAAAGGAGAGGGATGAGCTGATG	1799
Db	549	SerLysThrAspValGluGluLysMetTyrSerValThrLysGluArgAspAspLeuLys	568
QY	1800	GGTAAACTGAGGAGCGAAGAAAGGTCTGTGTGAACCTGAGCTGCAGTGTAGACTTACTA	1859
Db	569	AsnLysLeuLysAlaGluGluGluLysGlyAsnAspLeuLeuSerArgValAsnMetLeu	588
QY	1860	AAGAAGCGGCTTGATGGCATAGAGGAGGTAGAAAAGGGAAAATAAACCGAGGTAGG----	1913
Db	589	LysAsnArgLeuGlnSerLeuGluAlaIleGluLysAspPheLeuLysAsnLysLeuAsn	608
QY	1914	-----TCGTGCAAGGGGTCTGAGTTCCACTCCCTGCCCGGAGACAATAAGATCAGAGAACTA	1967
Db	609	GlnAspSerGlyLysSerThrThrAlaLeuHisGlnGluAsnAsnLysIleLysGluLeu	628

PS Claim 7; Page 68-72; 115pp; English.

XX The present sequence is the protein sequence of GIP130a, a novel 130 kDa

CC Goodpasture antigen binding protein (GPBP) interacting protein. GIP130a

CC cDNA was isolated from a skeletal muscle cDNA library. It represents an

CC mRNA species that results from faithful transcription and translation of

CC GIP90 genomic DNA. GIP90/130 polypeptides interact with GPBP and are

CC capable of aggregation. They can be used to modify GPBP-GIP90/130

CC interactions, to modify GP90/130 aggregation, and to modulate gene

CC expression. The invention provides GIP90/130 polypeptides, portions of

CC them, antibodies, nucleic acid sequences, expression vectors and host

CC cells, as well as methods for detecting GIP90/130 polypeptides and

CC nucleic acids, and methods for treating an autoimmune disease or cancer

CC by modifying the expression or activity of one or more GIP90/130

XX polypeptides

SQ Sequence 1135 AA;

Alignment Scores:

Pred. No.:	3.33e-157	Length:	1135
Score:	2444.50	Matches:	541
Percent Similarity:	64.31%	Conservative:	214
Best Local Similarity:	46.08%	Mismatches:	338
Query Match:	31.62%	Indels:	81
DB:	6	Gaps:	19

US-10-788-793-1 (1-4364) x ABR62251 (1-1135)

QY 75 ATGAGATCACGAAATCAAGTGGAGAAAGTTTCATCTAACGGGCATGTCTCCTGCCCCAAG 134

Db 1 MetArgSerArg-----GlySerAspThrGluGlySerAlaGlnLysLysPheProArg 18

QY 135 TCCTCCATCATCAGCAGTGATGGTGAAGGGCCCCCTCAGAAGATGCAAAAAAGAACAAAG 194

Db 19 HisThr-----LysGlyHisSerPheGlnGlyProLysAsnMet 31

QY 195 GCCAATCGGAAGGAGGAG---GATGTTCATGGCTTCGGAACTATCAAAAGGCACCTCAAA 251

Db 32 LysHisArgGlnGlnAspLysAspSerProSerGluSerAspVal-----IleLeu 48

QY 252 CCATCTGGAGAAAGTGAGAAA-----AAGACTAAGAAGTCTGTGGAGTTATCC 299

Db 49 ProCysProLysAlaGluLysProHisSerGlyAsnGlyHisGlnAlaGluAspLeuSer 68

QY 300 AAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGAGTTGCAGGTCGAGAAGAT 359

Db 69 ArgAspAspLeuLeuPheLeuLeuSerIleLeuGluGlyGluLeuGlnAlaArgAspGlu 88

QY 360 GTCATCCACATGCTGAGGACAGAGAAACCAAGCCCGAGGTTCTGGAGGCACACTATGGA 419

Db 89 ValIleGlyIleLeuLysAlaGluLysMetAspLeuAlaLeuLeuGluAlaGlnTyrGly 108

QY 420 TCTGCAGAACCTGAGAAAGTGCTTCGGGTCTCGCACCAGATGCCATCCTTGTCTCAAGAG 479

Db 109 PheValThrProLysLysValLeuGluAlaLeuGlnArgAspAlaPheGlnAlaLysSer 128

QY 480 AAGTCCATAGGAGACGCTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAA 539

Db 129 ThrProTrpGlnGluAspIleTyrGluLysProMetAsnGluLeuAspLysValValGlu 148

QY 540 AAGCAGAAGGAGACGTACCGCGCATGTCTAGAGCAGCTGTCTGGCTGAGAGTGTCTAC 599

Db 149 LysHisLysGluSerTyrArgArgIleLeuGlyGlnLeuLeuValAlaGluLysSerArg 168

QY 600 AGCGGCACCGTGTACGAGTGGAGAACGAGAAAGCACAGCACACTGACTACATGAACAAG 659

Db 169 ArgGlnThrIleLeuGluLeuGluGluLysArgLysHisLysGluTyrMetGluLys 188

QY 660 AGCGACGACTTCACCAACTGCTGGAGCAGGAGCGAGAGAGGTTGAAAAAGCTCCTTGA 719

Db 189 SerAspGluPheIleCysLeuLeuGluGlnGluCysGluArgLeuLysLysLeuIleAsp 208

QY 720 CAAGAAAAAGCTTTACCAAGCCCGCAAGAAAAAGGAAAAACGCTAAGCGGCTCAACAAACTT 779

Db 209 GlnGluIleLysSerGlnGluGluLysGluGlnGluLysGluLysArgValThrThrLeu 228

QY 780 CGAGATGAGCTTGTGAAGCTCAAGTCCCTTCGCCCTCATGTTGGTGGACGAGAGCAGATG 839

Db 229 LysGluGluLeuThrLysLeuLysSerPheAlaLeuMetValValAspGluGlnArg 248

QY 840 CACATCGAGCAACTGGCCCTGCAGAGTCAAGAAATCCAGGACCTCACTCAGAAAGCTGAGG 899

Db 249 LeuThrAlaGlnLeuThrLeuGlnArgGlnLysIleGlnGluLeuThrThrAsnAlaLys 268

QY 900 GAGGAGGAAGAAAACTCAAAGCGGTCACTTACAAATCCAAGGAAGACCGCCAGAAGCTG 959

Db 269 GluThrHisThrLysLeuAlaLeuAlaGluAlaArgValGlnGluGlnLysAla 288

QY 960 CTCAAAGTTAGAGTGGACTTCGAACACAAAGGCTCGAGGTTTTTCCCAGGAGACGAAGAG 1019

Db 289 ThrArgLeuGluLysGluLeuGlnThrGlnThrThrLysPheHisGlnAspGlnAspThr 308

QY 1020 ATGAACGCCAAATTTGGCGAATCAAGAAATCTCACAACCCGGCAACTTCGACTCAAACTGGTT 1079

Db 309 IleMetAlaLysLeuThrAsnGluAspSerGlnAsnArgGlnLeuGlnLysLeuAla 328

QY 1080 GGCTTATCGCAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCCTTCAGAAAGGCAGAG 1139

Db 329 AlaLeuSerArgGlnIleAspGluLeuGluGluThrAsnArgSerLeuArgLysAlaGlu 348

QY 1140 GAAGAGCTCCAGGAGCTGAGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAAATCCAGTCTC 1199

Db 349 GluGluLeuGlnAspIleLysGluLysIleSerLysGlyGluTyrGlyAsnAlaGlyIle 368

QY 1200 ATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTCTGAGATGGAGGGCAAGGATGAAGAG 1259

Db 369 MetAlaGluValGluGluLeuArgLysArgValLeuAspMetGluGlyLysAspGluGlu 388

QY 1260 ATCAGGAAGACCGAGGCCCAGTCCGGGAGTGAAGAAAGAGCTCCAAGAGAGAAACAC 1319

Db 389 LeuIleLysMetGluGluGlnCysArgAspLeuAsnLysArgLeuGluArgGluThrLeu 408

QY 1320 CACAGCAAGGAACCTTAGACTAGAAAGTGGAGAGCTGCAGAAGAGGATGTCTGAGCTGGAG 1379

Db 409 GlnSerLysAspPheLysLeuGluValGluLysLeuSerLysArgIleMetAlaLeuGlu 428

QY 1380 AAGCTGGAGGAACGCTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTG 1439

Db 429 LysLeuGluAspAlaPheAsnLysSerLysGlnGluCysTyrSerLeuLysCysAsnLeu 448

QY 1440 GAGAAGGAGAAACCTTAACCAAGACCTGTCTGAACGAGCTGGAGGTGGTCAAGAGTCGA 1499

Db 449 GluLysGluArgMetThrThrLysGlnLeuSerGlnGluLeuGluSerLeuLysValArg 468

QY 1500 GTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAAGCCGAGTTAAGCCTCAAGAT 1559

Db 469 IleLysGluLeuGluAlaIleGluSerArgLeuGluLysThrGluPheThrLeuLysGlu 488

QY 1560 GACCTTACAAAGCTGAAGTCCCTTCACCTGTGATGCTGGTGGATGAGAGGAAAAATATGATG 1619

Db 489 AspLeuThrLysLeuLysThrLeuThrValMetPheValAspGluArgLysThrMetSer 508

QY 1620 GAGAAAAATAAGCAAGAGAGAGAAAGTGGATGGGTGGAATAAAAACTTTAAGGTGGAG 1679

Db 509 GluLysLeuLysLysThrGluAspLysLeuGlnAlaAlaSerSerGlnLeuGlnValGlu 528

QY 1680 CAGGAAAAAGTCATGGATGTGACGGAAAAAGCTAATCGAGGAAAGCAAGAGCTTTAAAAA 1739

Db 529 GlnAsnLysValThrThrValThrGluLysLeuIleGluGluThrLysArgAlaLeuLys 548

QY 1740 CTCAAATCTGAAATGGAGAAAGAGTACAGTCTGACAAAGGAGAGGAGGATGAGCTGATG 1799

Db 549 SerLysThrAspValGluGluLysMetTyrSerValThrLysGluArgAspAspLeuLys 568

QY 1800 GGTAACACTGAGGACCGAAGAAAGGTCCTCTGTGAACCTGAGCTGCAGCTGTAGACTTACTA 1859

Db 569 AsnLysLeuLysAlaGluGluLysGlyAsnAspLeuLeuSerArgValAsnMetLeu 588
QY 1860 AAGAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGGGAAATAAACCGGAGGTAGG----- 1913
Db 589 LysAsnArgLeuGlnSerLeuGluAlaIleGluLysAspPheLeuLysAsnLysLeuAsn 608
QY 1914 -----TCGTGCAAGGGGTCTGAGTTCACTGCCCGGAAGACAATAAGATCAGAGAACTA 1967
Db 609 GlnAspSerGlyLysSerThrThrAlaLeuHisGlnGluAsnAsnLysIleLysGluLeu 628
QY 1968 ACGCTTGAAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTTGGAGGTGGTGGAGGGGAC 2027
Db 629 SerGlnGluValGluArgLeuLysLeuLysLeuLysAspMetLysAlaIleGluAspAsp 648
QY 2028 TTGATGAAGACCGAGGACGAATATGACCAAGTTGGAGCAGAAAGTTCAGAACCGCAGCAGGAT 2087
Db 649 LeuMetLysThrGluAspGluTyrGluThrLeuGluArgArgTyrAlaAsnGluArgAsp 668
QY 2088 AAGGCAAACTTCTCTCCAGCAGCTCGAGGAAATCAAAACCAAAATGGCCAAGCACAA 2147
Db 669 LysAlaGlnPheLeuSerLysGluLeuGluHisValLysMetGluLeuAlaLysTyrLys 688
QY 2148 GCCATAGAGAAAGGGGAGCCGTGAGCCAGGAAGCCGAACCTCGACACAGGTTTCGGCTG 2207
Db 689 LeuAlaGluLysThrGlu---ThrSerHisGluGlnTrpLeuPheLysArgLeuGlnGlu 707
QY 2208 GAGGAGGCTAAAGTCGTGATTATCAGGCCGAGGTGCAGGCTCTCAAGGAGAAGATCCAC 2267
Db 708 GluGluAlaLysSerGlyHisLeuSerArgGluValAspAlaLeuLysGluLysIleHis 727
QY 2268 GAGCTGATGAACAAGGAGACCACTGTCTCAGCTCCCAAGTCGACTATTTCGGTCTTCAG 2327
Db 728 GluTyrMetAlaThrGluAspLeuIleCysHisLeuGlnGlyAspHisSerValLeuGln 747
QY 2328 CAAAGATTATGGAAGAAGAACTAAGAACAAACATGGGAGGGAGGTCTCAATCTG 2387
Db 748 LysLysLeuAsnGlnGlnGluAsnArgAsnArgAspLeuGlyArgGluIleGluAsnLeu 767
QY 2388 ACCAAGGAGCTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGC 2447
Db 768 ThrLysGluLeuGluArgTyrArgHisPheSerLysSerLeuArgProSerLeuAsnGly 787
QY 2448 CGAAGGATGGTGACGTGCCTGTGGCTCCACTGGGGTGAGACCGGCGGTGTGCGGG 2507
Db 788 ArgArgIleSerAspProGlnValPheSerLysGluValGlnThrGluAlaVal----- 805
QY 2508 GATGCTGCGGAGGAGGAGACCCCG-----GCTGTG 2537
Db 806 -----AspAsnGluProProAspTyrLysSerLeuIleProLeuGluArgAlaVal 822
QY 2538 TTCATTGCAAAATCCTTCCAGGAG--GAAAAATCACATCATGAGTAATCTTCGACAGGTA 2594
Db 823 IleAsnGlyGlnLeuTyrGluGluSerGluAsnGlnAsp----- 835
QY 2595 GGCCTGAAGAAACCCATGGAACGGTCTCGGTCTCTC-----GACAGGTAT 2639
Db 836 -----GluAspProAsnAspGluGlySerValLeuSerPheLysCysSerGlnSerThr 853
QY 2640 CCCCCAGCAGCGAATGAGTCAACCATGACGAAGTCTTGGATTCTTGGATGAGAAAAGA 2699
Db 854 ProCysProValAsn-----ArgLysLeuTrpIleProTrpMetLysSerLys 869
QY 2700 GAA-----AACGGTCTCTCCACTCCGACGAGAGAAAGGCCAGGCCAAACCCAG 2747
Db 870 GluGlyHisLeuGlnAsnGly-----LysMetGlnThrLysProAsnAla 884
QY 2748 GGTGAGGGCACCCCGGGAGCTGGTCTTAGCACCAAGCAGGGCCAGCCCTACACATC 2807
Db 885 AsnPheValGlnProGlyAspLeuValLeuSerHisThrProGlyGlnProLeuHisIle 904
QY 2808 CGTGTGACACCATCATGAGAACAGCACTGCCACCTGGAGATCAACAAGCCCCACATCT 2867
Db 905 LysValThrProAspHisValGlnAsnThrAlaThrLeuGluIleThrSerProThrThr 924

QY 2868 GAA-----GAGTTTTTCTTAGTACCACCGTCACTTCCCTACCTTAGGCAACCCAGAAACCA 2921
Db 925 GluSerProHisSerTyrThrSerThrAlaValIleProAsnCysGlyThrProLysGln 944
QY 2922 AGAATAACCATATTCCATCACCCAATGTCATGTGCGAAAAGCCCCAAA---AGTCAGAT 2978
Db 945 ArgIleThrIleLeuGlnAsnAlaSerIleThrProValLysSerLysThrSerThrGlu 964
QY 2979 CCTACTCTCGGCCCCAGAACGAGCCATGTCCCCTGTCCAGATTACTATTTCAGAGAG 3038
Db 965 AspLeuMetAsnLeuGluGlnGlyMetSerProIleThrMetAlaThrPheAlaArgAla 984
QY 3039 AAGAGCCCGAAAGGTGGAAGGAGCGCCTTTGTCGACAGGCTGCATCCCCATCCAAATC 3098
Db 985 GlnThrProGluSerCysGlySerLeuThrProGluArgThrMetSerProIleGlnVal 1004
QY 3099 ATGACGGTGTCAACATCTGCAGCTCCCAGTGAATCGCTGTCTCTCTGAATCTCAGGAA 3158
Db 1005 LeuAlaValThrGlySerAlaSerSerProGluGlnGlyArgSerProGluProThrGlu 1024
QY 3159 GTGCCTATGGGAAGGACTATCTCAAAGTCACCCCGGAAAAACAAACTGTTCCAGCCCC 3218
Db 1025 IleSerAlaLysHisAlaIlePheArgValSerProAspArgGlnSerSerTrpGlnPhe 1044
QY 3219 GTGCGGAAGTACAACCTCAATGTCTAATATCATCACCGGAAGACAATAAAATTCACATT 3278
Db 1045 GlnArgSerAsnSerAsnSerSerValIleThrThrGluAspAsnLysIleHisIle 1064
QY 3279 CACCTGGTCTCTCAGTTTAAGCGATCTCTCTGGCCCTGCGCTGAAGGCGTGAGCCCAGTT 3338
Db 1065 HisLeuGlySerProTyrMetGlnAla-----ValAlaSerProValArgProAla 1081
QY 3339 ATCACCGTCCGGCTGTCAACGTGACAGCGGAGAGGAGGTTTCTACAGGCACAGTCCTT 3398
Db 1082 SerProSerAlaProLeuGlnAspAsnArgThrGlnGlyLeuIleAsnGlyAlaLeuAsn 1101
QY 3399 CGCTCTCCCAGGAACCACTCTCTTCAAGACCCGGTGCTAGCAAAAGTGACCAGCACTATA 3458
Db 1102 LysThr-----ThrAsnLysValThrSerSerIle 1111
QY 3459 ACTATAACCCCGGTCAAAACGTATCCACACGAGGAACCCCAA 3500
Db 1112 ThrIleThrProThrAlaThrProLeuProArgGlnSerGln 1125

RESULT 10
ABR69651
ID ABR69651 standard; protein; 893 AA.
XX
AC ABR69651;
XX 12-AUG-2003 (first entry)
XX Human CGDD-51 protein.
XX Human; cytostatic; antiarteriosclerotic; anticonvulsant; nootropic;
KW neuroprotective; cerebroprotective; anti-HIV; antiallergic;
KW antiinflammatory; gynaecological; cancer; atherosclerosis; epilepsy;
KW Huntington's disease; stroke; AIDS; allergy; placenta; reproductive;
KW CGDD; cell growth; cell differentiation; cell death.
OS Homo sapiens.
XX
PN WO2003027263-A2.
XX
PD 03-APR-2003.
XX
XX 26-SEP-2002; 2002WO-US031095.
PR 28-SEP-2001; 2001US-0326389P.
PR 05-OCT-2001; 2001US-0327380P.
PR 05-OCT-2001; 2001US-0328186P.
PR 12-OCT-2001; 2001US-0329690P.

PR 26-OCT-2001; 2001US-0345384P.
PR 26-OCT-2001; 2001US-0348165P.
PR 02-NOV-2001; 2001US-0350219P.
PR 09-NOV-2001; 2001US-0344518P.
PR 09-NOV-2001; 2001US-0345143P.
PR 16-NOV-2001; 2001US-0332375P.
PR 03-DEC-2001; 2001US-0336908P.
PR 07-DEC-2001; 2001US-0340747P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Azimzai Y, Baughn MR, Becha SD, Borowsky ML, Chawla NK;
PI Elliott VS, Emerling BM, Gandhi AR, Gietzen KJ, Gorvad AE;
PI Griffin JA, Hafalia AJA, Ison CH, Kable AE, Kalafus DP;
PI Lehr-Mason PM, Lu DAM, Marquis JP, Nguyen DB, Ramkumar J;
PI Richardson TW, Sapperstein SK, Swarnakar A, Tang YT, Tran UK;
PI Warren BA, Xu Y, Yao MG, Yue H, Yue H;
XX
DR WPI; 2003-421159/39.
DR N-PSDB; ACC90628.
XX
PT New human proteins associated with cell growth, differentiation, and
PT death (CGDD), useful for diagnosing, treating and preventing diseases or
PT conditions associated with the aberrant CGDD expression e.g. cancer,
PT AIDS, or epilepsy.
XX
PS Claim 1; Page 302-303; 350pp; English.
XX
CC The invention relates to an isolated polypeptide associated with cell
CC growth, differentiation and death (CGDD). Also disclosed are the
CC polynucleotides encoding the polypeptides. The polypeptides and
CC polynucleotides are useful in diagnosing, treating and preventing
CC diseases or conditions associated with the decreased expression or over
CC expression of CGDD. Such diseases include cell proliferative (e.g.
CC cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's
CC disease, stroke), immune/inflammatory (e.g. AIDS, allergies) and
CC reproductive disorders, or disorders of the placenta. They are also
CC useful in assessing the effects of exogenous compounds on the expression
CC of nucleic acid and amino acid sequences of CGDD. The CGDD or its
CC fragments are useful in screening compounds for effectiveness as an
CC agonist or antagonist of the polypeptides, or in altering the expression
CC of the target polynucleotide and compounds that specifically bind to or
CC modulate the activity of the polypeptide. Microarrays consisting
CC polynucleotides of the invention are useful in monitoring or measuring
CC protein-protein interactions, drug-target interactions, and gene
CC expression profiles. Sequences given in records AAR69601-AAR69657
CC represent CGDD polypeptides of the invention
XX
SQ Sequence 893 AA;

Alignment Scores:
Pred. No.: 5.18e-122 Length: 893
Score: 1926.00 Matches: 424
Percent Similarity: 64.21% Conservative: 168
Best Local Similarity: 45.99% Mismatches: 266
Query Match: 24.91% Indels: 64
DB: 6 Gaps: 14

US-10-788-793-1 (1-4364) x ABR69651 (1-893)

QY 816 ATGTTGGTGACGAGGAGGAGATGCACATCGAGCAACTGGCGCTGCAGAGTCAGAAAGTC 875
Db 1 MetValValAspGluGlnArgLeuThrAlaGlnLeuThrLeuGlnArgGlnLysIle 20

QY 876 CAGGACCTCACTCAGAAAGCTGAGGGAGGAGGAAGAAAACTCAAAGCGGTCACTTACAAA 935
Db 21 GlnGluLeuThrThrAsnAlaLysGluThrHisThrLysLeuAlaGluAlaArg 40

QY 936 TCCAAGGAAGACCGCCAGAAGCTGCTCAAGTTAGAAGTGGACTTCGAACACAAAGGCCTCG 995
Db 41 ValGlnGluGluGlnLysAlaThrArgLeuGluLysGluLeuGlnThrGlnThrThr 60

QY 996 AGGTTTTCACGAGCAGGAAGAGATGAACGCCAAATTGGCGAATCAAGATCTCACAAAC 1055

Db 61 LysPheHisGlnAspGlnAspThrIleMetAlaLysLeuThrAsnGluAspSerGlnAsn 80
QY 1056 CGGCAACTTCGACTCAAAAGCTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGAAGAGACC 1115
Db 81 ArgGlnLeuGlnGlnLysLeuAlaAlaLeuSerArgGlnIleAspGluLeuGluGluThr 100
QY 1116 AATAAAAGCCTTCAGAAAGGCAGAGGAGAGCTCCAGGAGCTGAGAGAGAAAATTGCCAAA 1175
Db 101 AsnArgSerLeuArgLysAlaGluGluGluLeuGlnAspIleLysGluLysIleSerLys 120
QY 1176 GGGGAATGTGGAAACTCCAGTCTCATGTCGGGGAAGTGGAGAGTCTGCGCAAGCGCGTGCTT 1235
Db 121 GlyGluTyrGlyAsnAlaGlyIleMetAlaGluValGluGluLeuArgLysArgValLeu 140
QY 1236 GAGATGGAGGCAAGGATGAAGAGATCACGAAGACCGAGGCCAGTGCCTGGGAGCTGAAG 1295
Db 141 AspMetGluGlyLysAspGluGluLeuIleLysMetGluGluGlnCysArgAspLeuAsn 160
QY 1296 AAGAAGCTCCAAGAGGAAGAACACACACAGCAAGGAACTTAGACTAGAAAGTGGAGAAGCTG 1355
Db 161 LysArgLeuGluArgGluThrLeuGlnSerLysAspPheLysLeuGluValGluLysLeu 180
QY 1356 CAGAAGAGGATGCTCGAGCTGGAGAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGA 1415
Db 181 SerLysArgIleMetAlaLeuGluLysLeuGluAspAlaPheAsnLysSerLysGlnGlu 200
QY 1416 TGCACCCAGCTCCATCTGAACCTGGAGAGGAGGAAGAACCTAACCAAGACCTGCTGAAC 1475
Db 201 CysTyrSerLeuLysCysAsnLeuGluLysGluArgMetThrThrLysGlnLeuSerGln 220
QY 1476 GAGCTGGAGGTCAAGAGTCAAGTTCGATTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAG 1535
Db 221 GluLeuGluSerLeuLysValArgIleLysGluLeuGluAlaIleGluSerArgLeuGlu 240
QY 1536 AAGGCCGAGTTAAGCTCAAAGATGACCTTACAAAGCTGAAGTCTTCACTGTGATGCTG 1595
Db 241 LysThrGluPheThrLeuLysGluAspLeuThrLysLeuLysThrLeuThrValMetPhe 260
QY 1596 GTGGATGAGAGGAAAAAATATGATGGAGAAAAATAAGCAAGAGAGGAAAGTGGATGGG 1655
Db 261 ValAspGluArgLysThrMetSerGluLysLeuLysLysThrGluAspLysLeuGlnAla 280
QY 1656 TTGAATAAAACTTTAAGGTGGAGCAGGGAAGATCATGGATGTGACCGGAAAGCTAATC 1715
Db 281 AlaSerSerGlnLeuGlnValGluGlnAsnLysValThrThrValThrGluLysLeuIle 300
QY 1716 GAGGAAAGCAAGAGCTTTTAAACTCAAATCTGAAATGGAGGAAAAGGATACAGTCTG 1775
Db 301 GluGluThrLysArgAlaLeuLysSerLysThrAspValGluGluLysMetTyrSerVal 320
QY 1776 ACAAGGAGAGGATGAGTGTGATGGTTAAACTGAGGAGCGAAGAAAGTCTGTGTGAA 1835
Db 321 ThrLysGluArgAspAspLeuLysAsnLysLysLeuLysAlaGluGluLysGlyAsnAsp 340
QY 1836 CTGAGCTGCAGTCTAGACTTACTAAAGAAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGG 1895
Db 341 LeuLeuSerArgValAsnMetLeuLysAsnArgLeuGlnSerLeuGluAlaIleGluLys 360
QY 1896 GAAATAAACCGAGGTAGG-----TCGTGCAAGGGTCTGAGTTCCACCTGCCCG 1943
Db 361 AspPheLeuLysAsnLysLeuAsnGlnAspSerGlyLysSerThrThrAlaLeuHisGln 380
QY 1944 GAAGACAATAAGATCAGAGAACTAACGCTTGAATCGAGAGACTGAAGAAACGGCTCCAG 2003
Db 381 GluAsnAsnLysIleLysGluLeuSerGlnGluValGluArgLeuLysLeuLysLeuLys 400
QY 2004 CAGTTGGAGGTGCTGGAGGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAG 2063
Db 401 AspMetLysAlaIleGluAspAspLeuMetLysThrGluAspGluTyrGluThrLeuGlu 420
QY 2064 CAGAAGTTCCAGAACCGAGCAGGATGAAGGCAAACTTCTCTCCAGCAGCTCGAGGAAATC 2123

Db 421 ArgArgTyrAlaAsnGluArgAspLysAlaGlnPheLeuSerLysGluLeuGluHisVal 440
QY 2124 AAACACCAAAATGCGCAAGCACAAAGCCATAGAGAAAGGGAGCGCGTGAGCCAGGAAGCC 2183
Db 441 LysMetGluLeuAlaLysTyrLysLeuAlaGluLysThrGlu--ThrSerHisGluGln 459
QY 2184 GAACTGCGACACAGGTTTCGGCTGGAGGAGGCTAAAAAGTCGTGATTTACAGGCGGAGGTG 2243
Db 460 TrpLeuPheLysArgLeuGluGluGluAlaLysSerGlyHisLeuSerArgGluVal 479
QY 2244 CAGGCTCTCAAGGAGAAAGATCCACGAGCTGATGAACAAGGAAGACACGAGCTGTCTCAGCTC 2303
Db 480 AspAlaLeuLysGluLysIleHisGluTyrMetAlaThrGluAspLeuIleCysHisLeu 499
QY 2304 CAAGTCGACTATTCGGTCCTTCAGCAAAAGATTTATGGAAGAGAAACTAAGAACAAAGAAC 2363
Db 500 GlnGlyAspHisSerValLeuGlnLysLysLeuAsnGlnGlnGluAsnArgAsnArgAsp 519
QY 2364 ATGGGGAGGAGGTCTCAATCTGACCAAGAGCTAGAGCTTCCAAGCGGTACAGCCGA 2423
Db 520 LeuGlyArgGluIleGluAsnLeuThrLysGluLeuGluArgTyrArgHisPheSerLys 539
QY 2424 GCTCTCAGGCGAGTGGGAACGGCCGAAGGATGGTGGACGTGCCTGTGGCCTCCACTGGG 2483
Db 540 SerLeuArgProSerLeuAsnGlyArgArgIleSerAspProGlnValPheSerLysGlu 559
QY 2484 GTGCAGACCGAGGCGGTGTGCGGGGATGTGCGGGAGGAGGAGACCCCG----- 2531
Db 560 ValGlnThrGluAlaVal-----AspAsnGluProProAspTyrLysSer 574
QY 2532 -----GCTGTGTTTCATTTCGCAAAATCCTTCCAGGAG---GAAAAATCAC 2570
Db 575 LeuIleProLeuGluArgAlaValIleAsnGlyGlnLeuTyrGluGluSerGluAsnGln 594
QY 2571 ATCATGAGTAATCTTCGACAGGTAGGCGCTGAAGAAACCCATGGAACGGTCCCTCGGTCCTC 2630
Db 595 Asp-----GluAspProAsnAspGluGlySerValLeu 605
QY 2631 -----GACAGGTATCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCT 2675
Db 606 SerPheLysCysSerGlnSerThrProCysProValAsn-----ArgLysLeu 621
QY 2676 TGGATTCCCTGGATGAGAAAAAGAGAA-----AACGGTCCTTCCACTCCGCAG 2723
Db 622 TrpIleProTrpMetLysSerLysGluGlyHisLeuGlnAsnGly----- 636
QY 2724 GAGAAAGGGCCAGGCCAACAGGGTGCAGGGGCACCCCGGGGAGCTGGTCCTAGCACCA 2783
Db 637 LysMetGlnThrLysProAsnAlaAsnPheValGlnProGlyAspLeuValLeuSerHis 656
QY 2784 AAGCAGGGCCAGCCCCTACACATCCGTGTGACACCAGATCATGAGAACACGACTGCCACC 2843
Db 657 ThrProGlyGlnProLeuHisIleLysValThrProAspHisValGlnAsnThrAlaThr 676
QY 2844 CTGGAGATCACAAGCCCCACATCTGAA-----GAGTTTTTCTCTAGTACCACCGTCAAT 2897
Db 677 LeuGluIleThrSerProThrThrGluSerProHisSerTyrThrSerThrAlaValIle 696
QY 2898 CCTACCTTAGGCAACCCAGAAACCAAGAATAACCATTAATTCATCACCCAATGTCATGTCG 2957
Db 697 ProAsnCysGlyThrProLysGlnArgIleThrIleLeuGlnAsnAlaSerIleThrPro 716
QY 2958 CAAAAGCCCAAA--AGTGCAGATCCTACTCTCGGCCCAAGAGCCCATGTCCCCTGTC 3014
Db 717 ValLysSerLysThrSerThrGluAspLeuMetAsnLeuGluGlnGlyMetSerProIle 736
QY 3015 ACGATTACTACTATTTCCAGAGAGAGAGCCCGGGAAGGTGGAAGGAGCGCCTTTGCCCGAC 3074
Db 737 ThrMetAlaThrPheAlaArgAlaGlnThrProGluSerCysGlySerLeuThrProGlu 756
QY 3075 AGGCCTGCATCCCCCATCCAAATCATACGGTGTCAACATCTGCAGTCTCCCACTGAAATC 3134
Db 757 ArgThrMetSerProIleGlnValLeuAlaValThrGlySerAlaSerSerProGluGln 776

QY 3135 GCTGTCTCTCTGAATCTCAGGAAGTGCCTATGGGAAGGACTATCTCAAAGTCACCCCG 3194
Db 777 GlyArgSerProGluProThrGluIleSerAlaLysHisAlaIlePheArgValSerPro 796
QY 3195 GAAAAACAACACTGTTCCAGCCCCCGTGGGAAAGTACAACTCCAATGTATAATCATCAC 3254
Db 797 AspArgGlnSerSerTrpGlnPheGlnArgSerAsnSerAsnSerSerSerValIleThr 816
QY 3255 ACGGAAGACAATAAAATTCACATTTCACCTGGGTTCTCAGTTTAAAGCGATCTCTGGGCGCT 3314
Db 817 ThrGluAspAsnLysIleHisIleHisLeuGlySerProTyrMetGlnAla----- 833
QY 3315 GCCGCTGAAGCGGTGAGCCAGTTATCACCGTCCGGCTGTGTCAAACGTGACAGCGGAGAAG 3374
Db 834 ValAlaSerProValArgProAlaSerProSerAlaProLeuGlnAspAsnArgThrGln 853
QY 3375 GAGGTTTCTACAGGCACAGTCCTTCGCTCTCCAGGAACCACTCTCTTCAAGACCCCGGT 3434
Db 854 GlyLeuIleAsnGlyAlaLeuAsnLysThr----- 863
QY 3435 GCTAGCAAAAGTGACCACTATATACTATAAACCCCGGTCAACACGTATCCACACGAGGA 3494
Db 864 ThrAsnLysValThrSerSerIleThrIleThrProThrAlaThrProLeuProArgGln 883
QY 3495 ACCCAA 3500
Db 884 SerGln 885
RESULT 11
ABR62250
ID ABR62250 standard; protein; 764 AA.
XX
AC ABR62250;
XX
DT 22-SEP-2003 (first entry)
XX
DE GPBP-interacting protein GIP90.
XX
KW Goodpasture antigen binding protein interacting protein; GPBP; GIP90;
KW human; transcription factor; autoimmune disease; cancer;
KW immunosuppressive; cytostatic.
XX
OS Homo sapiens.
XX
PN WO2003048193-A2.
XX
PD 12-JUN-2003.
XX
PF 05-DEC-2002; 2002WO-EP013802.
XX
PR 07-DEC-2001; 2001US-0338287P.
PR 20-MAY-2002; 2002US-0382004P.
XX
PA (SAUS/) SAUS J.
PA (REVE/) REVERT-ROS F.
XX
PI Saus J, Revert-Ros F;
XX
DR WPI; 2003-505281/47.
DR N-PSDB; ACC83927.
XX
PT New Goodpasture antigen binding protein-interacting 90 and 130 kDa
PT polypeptides, useful for diagnosing and/or treating disorders associated
PT with the GIP90/130 polypeptide, such as autoimmune disorders and/or
PT cancer.
XX
PS Claim 6; Page 59-62; 115pp; English.
XX
CC The present sequence is the protein sequence of GIP90, a novel 90 kDa
CC Goodpasture antigen binding protein (GPBP) interacting protein. GIP90
CC cDNA was isolated following a yeast two-hybrid screen for GPBP
CC interactive proteins and use of a partial clone to screen a human

CC skeletal muscle cDNA library. The GIP90 gene was localised to chromosome
CC 3 (3q12). Comparison of GIP90 cDNA and genomic sequences revealed the
CC existence of an adenine at position 2720 in the cDNA that was not present
CC in GIP90 genomic DNA. Another cDNA fragment was isolated that encoded a
CC 130 kDa protein (GIP130a, see ABR62251) and resulted from faithful
CC transcription and translation of the GIP90 genomic sequence (i.e. no
CC A2720). GIP90/130 polypeptides interact with GPBP and are capable of
CC aggregation. They can be used to modify GPBP-GIP90/130 interactions, to
CC modify GP90/130 aggregation, and to modulate gene expression. The
CC invention provides GIP90/130 polypeptides, portions of them, antibodies,
CC nucleic acid sequences, expression vectors and host cells, as well as
CC methods for detecting GIP90/130 polypeptides and nucleic acids, and
CC methods for treating an autoimmune disease or cancer by modifying the
CC expression or activity of one or more GIP90/130 polypeptides
xx
SQ Sequence 764 AA;

Alignment Scores:

Pred. No.: 2.23e-115 Length: 764
Score: 1828.00 Matches: 386
Percent Similarity: 69.55% Conservative: 144
Best Local Similarity: 50.66% Mismatches: 210
Query Match: 23.65% Indels: 22
DB: 6 Gaps: 7

US-10-788-793-1 (1-4364) x ABR62250 (1-764)

QY 75 ATGAGATCAGCAAAATCAAGGTGGAGAAAGTTTCATCTTAACGGGCATGTCTCCTGCCCAAG 134
|||||
Db 1 MetArgSerArg-----GlySerAspThrGluGlySerAlaGlnLysLysPheProArg 18

QY 135 TCCTCCATCATCAGCAGTGCATGGTGTGAAGGGCCCTCAGAAAGATGCAGAAAGAACAAAG 194
|||
Db 19 HisThr-----LysGlyHisSerPheGlnGlyProLysAsnMet 31

QY 195 GCCAATCGGAAGGAGGAG---GATGTCATGGCTTCGGAACTATCAAAAGGCACCTCAAA 251
:::|||||
Db 32 LysHisArgGlnGlnAspLysAspSerProSerGluSerAspVal-----IleLeu 48

QY 252 CCATCTGGAGAAAGTGAGAAA-----AAGACTAAGAAGTCTGTGGAGTTATCC 299
|||
Db 49 ProCysProLysAlaGluLysProHisSerGlyAsnGlyHisGlnAlaGluAspLeuSer 68

QY 300 AAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGGAGTTGCAGGCTCGAGAAGAT 359
:::|||||
Db 69 ArgAspAspLeuLeuPheLeuLeuSerIleLeuGluGlyGluLeuGlnAlaArgAspGlu 88

QY 360 GTCATCCACATGCTGAGCACAGAGAAAACCAAGCCGAGGTTCTGGAGGCACACTATGGA 419
|||||
Db 89 ValIleGlyIleLeuLysAlaGluLysMetAspLeuAlaLeuLeuGluAlaGlnTyrGly 108

QY 420 TCTGCAGAACCTGAGAAAGTGTCTTCGGGTCTGCACCGAGATGCCATCCTTGTCTCAAGAG 479
|||||
Db 109 PheValThrProLysLysValLeuGluAlaLeuGlnArgAspAlaPheGlnAlaLysSer 128

QY 480 AAGTCCATAGGAGAACGCTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAA 539
|||||
Db 129 ThrProTrpGlnGluAspIleTyrGluLysProMetAsnGluLeuAspLysValValGlu 148

QY 540 AAGCAGAAGGAGACGTACCGCCGCATGCTAGAGCAGCTGTGTGGCTGGAGAGTGTAC 599
|||||
Db 149 LysHisLysGluSerTyrArgArgIleLeuGluLysArgLysHisLysGlnLeuValAlaGluLysSerArg 168

QY 600 AGCGGCACCGTGTACGAGCTGGAGAACGAGAGCAAGCACACTGACTACATGAACAAG 659
|||||
Db 169 ArgGlnThrIleLeuGluLeuGluGluGluLysArgLysHisLysGlnTyrMetGluLys 188

QY 660 AGCGACGACTTCACCAACCTGTCTGGAGCAGGAGCAGAGAGGTTGAAAAAGCTCCTGAA 719
|||||
Db 189 SerAspGluPheIleCysLeuLeuGluGlnGluCysGluArgLeuLysLysLeuIleAsp 208

QY 720 CAAGAAAAAGCTTACCAAGCCCGCAAGAAAAGGAAAAACGCTAAGCGGCTCAACAACTT 779
|||||

Db 209 GlnGluIleLysSerGlnGluGluLysGluGlnGluLysGluLysArgValThrThrLeu 228

QY 780 CGAGATGAGCTTGTGAAGCTCAAGTCCCTTCGCCCTCATGTTGGTGACGAGAGCAGATG 839
:::|||||
Db 229 LysGluGluLeuThrLysLeuLysSerPheAlaLeuMetValValAspGluGlnGlnArg 248

QY 840 CACATCGAGCAACTGGGCCCTGCAGAGTGCAGAACTCAGAAAGTCCAGGACCTCAGTCAGAGCTGAGG 899
|||||
Db 249 LeuThrAlaGlnLeuThrLeuGlnArgGlnLysIleGlnGluLeuThrThrAsnAlaLys 268

QY 900 GAGGAGGAAGAAAACTCAAGCGGTCACTTACAAAATCCAAGGAAGACCGCCAGAAAGCTG 959
|||
Db 269 GluThrHisThrLysLeuAlaLeuAlaGluAlaArgValGlnGluGluGlnLysAla 288

QY 960 CTCAGTTAGAGTGGACTTCGAACACCAAGGCCCTCGAGGTTTCCCAGGAGCAGCAAGAG 1019
:::|||||
Db 289 ThrArgLeuGluLysGluLeuGlnThrGlnThrThrLysPheHisGlnAspGlnAspThr 308

QY 1020 ATGAACGCCAAATTGGCGAATCAAGAAATCTCACAAACCGGCAACTTCGACTCAAACTGGTT 1079
:::|||||
Db 309 IleMetAlaLysLeuThrAsnGluAspSerGlnAsnArgGlnLeuGlnGlnLysLeuAla 328

QY 1080 GGCTTATCGAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAGGCAGAG 1139
|||||
Db 329 AlaLeuSerArgGlnIleAspGluLeuGluThrAsnArgSerLeuArgLysAlaGlu 348

QY 1140 GAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAAGGGGAATGTGGAACCTCCAGTCTC 1199
|||||
Db 349 GluGluLeuGlnAspIleLysGluLysIleSerLysGlyGluTyrGlyAsnAlaGlyIle 368

QY 1200 ATGGCGGAAGTGGAGAGTCTGCGCAAGCCGCTGTGAGATGGAGGGCAAGGATGAAGAG 1259
|||||
Db 369 MetAlaGluValGluLeuArgLysArgValLeuAspMetGluGlyLysAspGluGlu 388

QY 1260 ATCAGGAAGACCGAGGCCAGTCCGGGAGCTGAAGAAAGAGCTCCAAGAGGAAGAACAC 1319
:::|||||
Db 389 LeuIleLysMetGluGluGlnCysArgAspLeuAsnLysArgLeuGluArgGluThrLeu 408

QY 1320 CACAGCAAGGAACCTTAGACTAGAAGTGGAGAAGCTGCAGAAAGAGGATGTCTGAGCTGGAG 1379
|||||
Db 409 GlnSerLysAspPheLysLeuGluValGluLysLeuSerLysArgIleMetAlaLeuGlu 428

QY 1380 AAGCTGGAGGAAGCGTTTCAGCCGGAGTAAAGTCGGAATGCACCCAGCTCCATCTGAACCTG 1439
|||||
Db 429 LysLeuGluAspAlaPheAsnLysSerLysGlnGluCysTyrSerLeuLysCysAsnLeu 448

QY 1440 GAGAAGGAGAGAACCTTAACCAAGACCTTGTGAAACGAGCTGGAGGTGGTCAAGAGTCGA 1499
|||||
Db 449 GluLysGluArgMetThrThrLysGlnLeuSerGlnGluLeuGluSerLeuLysValArg 468

QY 1500 GTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAAGCCGAGTTAAGCCTCAAAGAT 1559
:::|||||
Db 469 IleLysGluLeuGluAlaIleGluSerArgLeuGluLysThrGluPheThrLeuLysGlu 488

QY 1560 GACCTTACAAAGCTGAAGTCTTCACTGTGATGCTGGTGGATGAGAGGAAAAATATGATG 1619
|||||
Db 489 AspLeuThrLysLeuLysThrLeuThrValMetPheValAspGluArgLysThrMetSer 508

QY 1620 GAGAAAAATAAGCAAGAGAGAGGAAAGTGGATGGTGTGAATAAAAACTTTAAGGTGGAG 1679
|||||
Db 509 GluLysLeuLysLysThrGluAspLysLeuGlnAlaAlaSerSerGlnLeuGlnValGlu 528

QY 1680 CAGGGAAGAAAGTCATGGATGTGACGGAAAAGCTTAATCGAGGAAAGCAAGAGCTTTTAAAA 1739
|||
Db 529 GlnAsnLysValThrThrValThrGluLysLeuIleGluGluThrLysArgAlaLeuLys 548

QY 1740 CTCAAATCTGAAATGGAGGAAAAGGAGTACAGTCTGACAAAGGAGGAGGATGAGCTGATG 1799
|||||
Db 549 SerLysThrAspValGluGluLysMetTyrSerValThrLysGluArgAspAspLeuLys 568

QY 1800 GGTAACCTGAGAGCGAAGAAAGAAAGGTCTCTGTGAAGCTGAGCTGCAGTGTAGACTTACTA 1859
|||||
Db 569 AsnLysLeuLysAlaGluGluLysGlyAsnAspLeuLeuSerArgValAsnMetLeu 588

XX Baak J, Mutter GL;
XX WPI; 2002-180084/23.
DR N-PSDB; ABK35549.
XX

PT Diagnosing breast cancer comprises determining expression of nucleic acid
PT molecules or expression products that are differentially expressed in
PT normal and malignant tissue.

XX Claim 37; Page 141-143; 219pp; English.

XX The present invention relates to a method for diagnosing breast cancer in
CC a subject suspected of having endometrial cancer. The method comprises
CC determining the expression of a set of human genes or expression products
CC in an endometrial sample suspected of being cancerous. The human genes of
CC the invention are differentially expressed in breast tumours
CC characterised as high or low MAI (mitotic activity index). These sets of
CC genes can be used to discriminate between high and low MAI breast
CC tumours. The invention also provides DNA and protein microarrays for
CC analysing the expression of the human genes and their protein products.
CC The methods and arrays are useful for the diagnosis and prognosis of
CC endometrial cancer, selecting and monitoring treatment regimes, and
CC identification of compounds useful for the treatment of endometrial
CC cancer. AAU84311-AAU84361 represent the human proteins of the invention
CC that are differentially expressed in breast cancer tissue

XX Sequence 752 AA;

Alignment Scores: Length: 752
Pred. No.: 5.57e-103 Matches: 363
Score: 1645.50
Percent Similarity: 64.16% Conservative: 140
Best Local Similarity: 46.30% Mismatches: 216
Query Match: 21.28% Indels: 65
DB: 5 Gaps: 13

US-10-788-793-1 (1-4364) x AAU84329 (1-752)

Qy 816 ATGTTGGTCGACGAGGAGGATGCATCGACCAACTGGCCCTGCAGAGTCAGAAAGTC 875
Db 1 MetValValAspGluGlnArgLeuThrAlaGlnLeuThrLeuGlnArgGlnLysIle 20
Qy 876 CAGGACCTCACTCAGAAAGCTGAGGAGGAGGAAGAAAACCTCAAAGCGGTCACTTACAA 935
Db 21 GlnGluLeuThrThrAsnAlaLysGluThrHisThrLysLeuAlaLeuAlaGluAlaArg 40
Qy 936 TCCAAGGAGACCGCCAGAGCTGCTCAAGTTAGAAAGTGGACTTCGAACACAAAGCCCTCG 995
Db 41 ValGlnGluGluGluGlnLysAlaThrArgLeuGluLysLeuGlnThrThr 60
Qy 996 AGGTTTCCAGGAGCACGAAAGATGAACGCCAAATTGGCGAATCAAGAATCTCACAA 1055
Db 61 LysPheHisGlnAspGlnAspThrIleMetAlaLysLeuThrAsnGluAspSerGlnAsn 80
Qy 1056 CGGCAACTTCGACTCAAACCTGGTTGGCTTATFCGCAAAAGGATTGAGGAGCTGGAAAGACC 1115
Db 81 ArgGlnLeuGlnGlnLysLeuAlaAlaLeuSerArgGlnIleAspGluLeuGluGluThr 100
Qy 1116 AATAAAAGCCTTCAGAAAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAATTGCCAAA 1175
Db 101 AsnArgSerLeuArgLysAlaGluGluGlnLeuGlnAspIleLysGluLysIleSerLys 120
Qy 1176 GGGGAATGTGAAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGGTGCTT 1235
Db 121 GlyGluTyrGlyAsnAlaGlyIleMetAlaGluVal----- 132
Qy 1236 GAGATGGAGGCAAGGATGAAGAGATCACGAAGACCGAGGCCAGTGCCTGGAGCTGAAG 1295
Db 133 -----GluGluLeuIleLysMetGluGluGlnCysArgAspLeuAsn 146
Qy 1296 AAGAAGCTCAAAGAGGAAGAACACCAACGCAAGGAACCTTAGACTAGAAAGTGGAGAAAGCTG 1355

Db 147 LysArgLeuGluArgGluThrLeuGlnSerLysAspPheLysLeuGluValGluLysLeu 166
Qy 1356 CAGAAAGAGGATGTCTGAGCTGGAGAAGCTGGAGGAAGCGTTACGCCGAGTAAGTCGGAA 1415
Db 167 SerLysArgIleMetAlaLeuGluLysLeuGluAspAlaPheAsnLysSerLysGlnGlu 186
Qy 1416 TGCACCCAGCTCCATCTGAACCTGGAGAAAGGAGAAACCTAAACCAAAGACCTGCTGAAC 1475
Db 187 CysTyrSerLeuLysCysAsnLeuGluLysGluArgMetThrThrLysGlnLeuSerGln 206
Qy 1476 GAGCTGGAGGTGGTCAACAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAG 1535
Db 207 GluLeuGluSerLeuLysValArgIleLysGluLeuGluAlaIleGluSerArgLeuGlu 226
Qy 1536 AAGGCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTGAAGTCTTCACTGTGATGCTG 1595
Db 227 LysThrGluPheThrLeuLysGluAspLeuThrLysLeuLysThrLeuThrValMetPhe 246
Qy 1596 GTGGATGAGAGGAAAAATATGATGGAGAAAAATAAGCAAGAGAGAGGAAAGTGGATGGG 1655
Db 247 ValAspGluArgLysThrMetSerGluLysLeuLysLysThrGluAspLysLeuGlnAla 266
Qy 1656 TTGAATAAAACCTTAAAGTGGAGCAGGGAAGATCATGGATGTGACGGAAAAGCTAATC 1715
Db 267 AlaSerSerGlnLeuGlnValGluGlnAsnLysValThrThrValThrGluLysLeuIle 286
Qy 1716 GAGGAAAGCAAGAGCTTTTAAAACTCAAATCTGAAATGGAGGAAAAAGGAGTACAGTCTG 1775
Db 287 GluGluThrLysArgAlaLeuLysSerLysThrAspValGluGluLysMetTyrSerVal 306
Qy 1776 ACAAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGGAAGAAAGGTCCTGTGAA 1835
Db 307 ThrLysGluArgAspAspLeuLysAsnLysLeuLysAlaGluGluGluLysGlyAsnAsp 326
Qy 1836 CTGAGCTGCAGTGTAGACTTACTAAAGAACGGCTTGTGTCATAGAGGAGGTAGAAAGG 1895
Db 327 LeuLeuSerArgValAsnMetLeuLysAsnArgLeuGlnSerLeuGluAlaIleGluLys 346
Qy 1896 GAAATAAACCGAGGTAGG-----TCGTGCAAGGGGTCTGAGTTACCTGCCCG 1943
Db 347 AspPheLeuLysAsnLysLeuAsnGlnAspSerGlyLysSerThrThrAlaLeuHisGln 366
Qy 1944 GAAGACAATAAGATCAGAGAACTAACGCTTGAATCGAGAGACTGAAGAAACGGCTCCAG 2003
Db 367 GluAsnAsnLysIleLysGluLeuSerGlnGluValGluArgLeuLysLeuLys 386
Qy 2004 CAGTTGGAGTGGTGGAGGGGACTTGATGAAGACCAGGACGAATATGACCAGTTGGAG 2063
Db 387 AspMetLysAlaIleGluAspAspLeuMetLysThrGluAspGluTyrGluThrLeuGlu 406
Qy 2064 CAGAAAGTTCAGAAACCGCAGCAGGATAAGGCAAACTTCTCTCCCAGCAGCTCGAGGAAATC 2123
Db 407 ArgArgTyrAlaAsnGluArgAspLysAlaGlnPheLeuSerLysGluLeuGluHisVal 426
Qy 2124 AAACACCAAAATGGCCCAAGCACAAAGCCCATAGAGAAAGGGAGGCCGTGAGCCAGGAAGCC 2183
Db 427 LysMetGluLeuAlaLysTyrLysLeuAlaGluLysThrGlu---ThrSerHisGluGln 445
Qy 2184 GAACTGCGACACAGGTTTCGGCTGGAGGAGGCTAAAGTCGTGATTTACAGGCCGAGGTG 2243
Db 446 TrpLeuPheLysArgLeuGlnGluGluAlaLysSerGlyHisLeuSerArgGluVal 465
Qy 2244 CAGGCTCTCAAGGAGAAGATCCACGAGCTGATGAACAAGGAAAGACCAAGCTGTCTCAGCTC 2303
Db 466 AspAlaLeuLysGluLysIleHisGluTyrMetAlaThrGluAspLeuIleCysHisLeu 485
Qy 2304 CAAGTCGACTATTCCGTCTTCAGCAAAGATTATGGAAGAAAGAAATAAGAACAAAGAAC 2363
Db 486 GlnGlyAspHisSerValCysLysLysLysLeuAsnGlnGlnGluAsnArgAsnArgAsp 505
Qy 2364 ATGGGGAGGAGGTCTCTCAATCTTGACCAAGGAGCTAGAGCTTCCAAAGCGCTACAGCCGA 2423
Db 506 LeuGlyArgGluIleGluAsnLeuThrLysGluLeuGluArgTyrArgHisPheSerLys 525

Db 121 GlyGluTyrGlyAsnAlaGlyIleMetAlaGluVal----- 132

QY 1236 GAGATGGAGGGCAAGGATGAAGAGATCATCAGAAAGACCGAGGCCAGTCGCCGGAGCTGAAG 1295

Db 133 -----GluGluLeuIleLysMetGluGluGlnCysArgAspLeuAsn 146

QY 1296 AAGAAAGCTCCAAGAGGAAGAAACACACAGCAAGAACTTAGACTAGAAAGTGGAGAAGCTG 1355

Db 147 LysArgLeuGluArgGluThrLeuGlnSerLysAspPheLysLeuGluValGluLysLeu 166

QY 1356 CAGAAGAGGATGCTGAGCTGGAGAAAGCTGGAGAAAGCGTTTCAGCCGGAGTAAGTCGGAA 1415

Db 167 SerLysArgIleMetAlaLeuGluLysLeuGluAspAlaPheAsnLysSerLysGlnGlu 186

QY 1416 TGCACCCAGCTCCATCTGAACCTGGAGAAAGGAGAAACCTTAACCAAGACCTGCTGAAC 1475

Db 187 CysTyrSerLeuLysCysAsnLeuGluLysGluArgMetThrThrLysGlnLeuSerGln 206

QY 1476 GAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAG 1535

Db 207 GluLeuGluSerLeuLysValArgIleLysGluLeuGluAlaIleGluSerArgLeuGlu 226

QY 1536 AAGCCCGAGTTAAGCCTCAAGATGACCTTACAAAGCTGAATCCTTCACTGTGATGCTG 1595

Db 227 LysThrGluPheThrLeuLysGluAspLeuThrLysLeuLysThrLeuThrValMetPhe 246

QY 1596 GTGGATGAGAGGAAAAATATGATGGAGAAAATAAAGCAAGAGAGAGGAAAGTGGATGGG 1655

Db 247 ValAspGluArgLysThrMetSerGluLysLeuLysLysThrGluAspLysLeuGlnAla 266

QY 1656 TTGAATAAAAACTTTAAGTGGAGCAGGGGAAAAGTCATGGATGTGACGGAAAAAGCTAATC 1715

Db 267 AlaSerSerGlnLeuGlnValGluGlnAsnLysValThrThrValThrGluLysLeuIle 286

QY 1716 GAGGAAAGCAAGAAGCTTTTAAAACTCAATCTGAAATGGAGGAAAAAGAGTACAGTCTG 1775

Db 287 GluGluThrLysArgAlaLeuLysSerLysThrAspValGluGluLysMetTyrSerVal 306

QY 1776 ACAAAGGAGAGGGATGAGCTGATGGGTAAACTGAGGAGCGAAGAAAGAGTCTGTGAA 1835

Db 307 ThrLysGluArgAspAspLeuLysAsnLysLeuLysAlaGluGluLysGlyAsnAsp 326

QY 1836 CTGAGCTGAGTGTAGACTTACTAAAGACGGCTTGATGGCATAGAGAGGTAGAAAGG 1895

Db 327 LeuLeuSerArgValAsnMetLeuLysAsnArgLeuGlnSerLeuGluAlaIleGluLys 346

QY 1896 GAAATAAACCGAGGTAGG-----TCGTGCAAGGGGTCTGAGTTCACCTGCCCG 1943

Db 347 AspPheLeuLysAsnLysLeuAsnGlnAspSerGlyLysSerThrThrAlaLeuHisGln 366

QY 1944 GAAGACATAAAGATCAGAGAACTAACCGTTGAAATCGAGAGACTGAAGAAAAACGGCTCCAG 2003

Db 367 GluAsnAsnLysIleLysGluLeuSerGlnGluValGluArgLeuLysLeuLysLeuLys 386

QY 2004 CAGTTGGAGGTGGAGGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAG 2063

Db 387 AspMetLysAlaIleGluAspAspLeuMetLysThrGluAspGluTyrGluThrLeuGlu 406

QY 2064 CAGAAGTTCCAGAACCGAGCAGGATAAGSCAAACTTCCTCTCCAGCAGCTCGAGGAATC 2123

Db 407 ArgArgTyrAlaAsnGluArgAspLysAlaGlnPheLeuSerLysGluLeuGluHisVal 426

QY 2124 AAACACCAAATGGCCCAAGCACAAAGCCATAGAGAAAGGGGAGGCCGTGAGCCAGGAAGCC 2183

Db 427 LysMetGluLeuAlaLysTyrLysLeuAlaGluLysThrGlu---ThrSerHisGluGln 445

QY 2184 GAACTCGACACAGGTTTCGGCTGGAGGAGCTAAAGTCGTGATTTACAGGCCGAGGTG 2243

Db 446 TrpLeuPheLysArgLeuGlnGluGluAlaLysSerGlyHisLeuSerArgGluVal 465

QY 2244 CAGGCTCTCAAGGAGAGAGATCCACGAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTC 2303

Db 466 AspAlaLeuLysGluLysIleHisGluTyrMetAlaThrGluAspLeuIleCysHisLeu 485

QY 2304 CAAGTCGACTATTTCGGTCTTTCAGCAAAAGATTTATTGAAGAAGAAACTAAGACAAGAAC 2363

Db 486 GlnGlyAspHisSerValCysLysLysLysLeuAsnGlnGlnGluAsnArgAsnArgAsp 505

QY 2364 ATGGGGAGGAGGTCTCTCAATCTGACCAAGGAGCTAGAGCTTTCAAAGCGCTACAGCCGA 2423

Db 506 LeuGlyArgGluIleGluAsnLeuThrLysGluLeuGluArgTyrArgHisPheSerLys 525

QY 2424 GCTCTCAGGCCGAGTGGGAACGGCCGGAAGGATGGTGGAGCTGCTGTGGCTCCACTGGG 2483

Db 526 SerLeuArgProSerLeuAsnGlyArgArgIleSerAspProGlnValPheSerLysGlu 545

QY 2484 GTGCAGACCGAGCGGTGTGCGGGGATGCTGCGGAGGAGGAGACCCCG----- 2531

Db 546 ValGlnThrGluAlaVal-----AspAsnGluProProAspTyrLysSer 560

QY 2532 -----GCTGTGTTTCATTCGAAATCCTTCCAGGAG---GAAAAATCAC 2570

Db 561 LeuIleProLeuGluArgAlaValIleAsnGlyGlnLeuTyrGluGluSerGluAsnGln 580

QY 2571 ATCATGAGTAATCTTCGACAGGTAGGCTGAAGAAACCCATGGAAACGGTCTCGGTCTC 2630

Db 581 Asp-----GluAspProAsnAspGluGlySerValLeu 591

QY 2631 -----GACAGGTATCCCCCAGCAGCGCAATGAGCTCACCATGAGGAAGTCT 2675

Db 592 SerPheLysCysSerGlnSerThrProCysProValAsn-----ArgLysLeu 607

QY 2676 TGGATTCCTTGGATGAGAAAAAGAGAA-----AACGGTCTTTCCTCCACTCCGCAG 2723

Db 608 TrpIleProTrpMetLysSerLysGluGlyHisLeuGlnAsnGly----- 622

QY 2724 GAGAAAGGGCCCCAGGCCAAACCAGGGTGCAGGGCACCCCGGGAGCTGGTCTTAGCACCA 2783

Db 623 LysMetGlnThrLysProAsnAlaAsnPheValGlnProGlyAspLeuValLeuSerHis 642

QY 2784 AAGCAGGGCCAGCCCTACACATCCGTGTGACACCATCATGAGAAACAGCAGCACTGCCACC 2843

Db 643 ThrProGlyGlnProLeuHisIleLysValThrProAspHisValGlnAsnThrAlaThr 662

QY 2844 CTGGAGATCACAAAGCCCCACATCTGAA-----GAGTTTCTCTAGTACCAACCGTCATT 2897

Db 663 LeuGluIleThrSerProThrThrGluSerProHisSerTyrThrSerThrAlaValIle 682

QY 2898 CCTACCTTAGGCAACCCAGAAACCAGAAATAACCATTATTCATCATCCCAATGTCTATGTCG 2957

Db 683 ProAsnCysGlyThrProLysGlnArgIleThrIleLeuGlnAsnAlaSerIleThrPro 702

QY 2958 CAAAAGCCCCAAA---AGTGCAGATCCTACTCTCGGCCCCAGAACGAGCCATGTCCCTGTTC 3014

Db 703 ValLysSerLysThrSerThrGluAspLeuMetAsnLeuGluGlnGlyMetSerProIle 722

QY 3015 ACGATTACTACTATTTCAGAGAGAAAGACCCCGGAAGGTGGAAAGGAGCGCCTTTGCCGAC 3074

Db 723 ThrMetAlaThrPheAlaArgAlaGlnThrProGluSerCysGlySerLeuThrProGlu 742

QY 3075 AGGCCTGCATCC 3086

Db 743 ArgThrMetSer 746

RESULT 15

AAM41802

ID AAM41802 standard; protein; 233 AA.

XX AC AAM41802;

XX DT 22-OCT-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 6733.

XX KW Human; neutropic; immunosuppressant; cytostatic; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI60958.
XX

Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries.

Example 2; SEQ ID NO 6733; 10078pp; English.

The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAM38642-AAM42213) with nootropic, immunosuppressant and cyostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders. Note: The sequence data for this patent did not form part of the printed specification

Sequence 233 AA;

Alignment Scores:
Pred. No.: 1.48e-63 Length: 233
Score: 1062.50 Matches: 211
Percent Similarity: 95.28% Conservative: 11
Best Local Similarity: 90.56% Mismatches: 10
Query Match: 13.74% Indels: 1
DB: 4 Gaps: 1

US-10-788-793-1 (1-4364) x AAM41802 (1-233)

QY 54 AGGAGTCGACACAGGTGGGAATGAGATCAGCAATCAAGGTGGAGAAAGTTCTTAAC 113
Db 1 ArgSerCysArgGlnValGlyMetArgSerArgAsnGlnGlyGlyGluSerAlaSerAsp 20
QY 114 GGGCATGTCTCCCAAGTCCTCCATCATCAGCAGTGATGGTGGTAAGGCCCTCA 173
Db 21 GlyHisIleSerCysProLysProSerIleIleGlyAsnAlaGlyGluLysSerLeuSer 40

QY 174 GAAGATGCA---AAAAAGAACAAAGCCCAATCGAAGAGGAGGATGTCTATGGCTTCGGA 230
Db 41 GluAspAlaLysLysLysLysLysSerAsnArgLysGluAspValMetAlaSerGly 60
QY 231 ACTATCAAAAGGCACCTCAAAACCATCTGGAGAAAGTGAAGAAAGACTAAGAAAGTCTGTG 290
Db 61 ThrValLysArgHisLeuLysThrSerGlyGluCysGluArgLysThrLysLysSerLeu 80
QY 291 GAGTTATCCAAGGAGGACCTCATCCAGCTCCCTGAGTATCATGGAAGGGGAGTTGAGGCT 350
Db 81 GluLeuSerLysGluAspLeuIleGlnLeuLeuSerIleMetGluGlyGluLeuGlnAla 100
QY 351 CGAGAAGATGTCATCCACATGCTGAGGACAGAGAAAAACCAAGCCCGAGGTTCTGGAGGCA 410
Db 101 ArgGluAspValIleHisMetLeuLysThrGluLysThrLysProGluValLeuGluAla 120
QY 411 CACTATGGATCTGCAGAACCTTGAGAAAGTGTCTCGGGTCTGCACCGAGATGCCATCCTT 470
Db 121 HisTyrGlySerAlaGluProGluLysValLeuArgValLeuHisArgAspAlaIleLeu 140
QY 471 GCTCAAGAGAAGTCCATAGGAGAAGACGTCTATGAGAAACCTATCTCAGAGCTGGACAGA 530
Db 141 AlaGlnGluLysSerIleGlyGluAspValTyrGluLysProIleSerGluLeuAspArg 160
QY 531 CTGGAGGAAAAAGCAGAGGAGACGTACCCGCGCATGCTAGAGCAGCTGCTGCTGCTGAG 590
Db 161 LeuGluGluLysGlnLysGluThrTyrArgArgMetLeuGluGlnLeuLeuAlaGlu 180
QY 591 AAGTGTCAAGCGCCGACCGTGTACGAGCTGGAGAACGAGAACGACACACTGACTAC 650
Db 181 LysCysHisArgArgThrValTyrGluLeuGluAsnGluLysHisLysHisThrAspTyr 200
QY 651 ATGAACAAGAGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGAGAGAGGTTGAAAAAG 710
Db 201 MetAsnLysSerAspAspPheThrAsnLeuLeuGluGlnGluArgGluArgLeuLysLys 220
QY 711 CTCCTTGAACAAGAAAAAGCTTACCAAGCCCGCAAGAA 749
Db 221 LeuLeuGluGlnGluLysAlaTyrGlnAlaArgLysGlu 233

Search completed: September 7, 2004, 13:44:37
Job time : 525 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 7, 2004, 14:02:41 ; Search time 443.5 Seconds
(without alignments)
6200.425 Million cell updates/sec

Title: US-10-788-793-1
Perfect score: 7731
Sequence: 1 ccactgggtcttcaaggga.....aaaaaaaaaaaaaaaaaaaa 4364

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1298764 seqs, 315065143 residues

Total number of hits satisfying chosen parameters: 2597528

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10788793/runat_07092004_134024_10911/app_query.fasta_1.4551
-DB=Published Applications AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10788793@cgn_1_1_714@runat_07092004_134024_10911
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

* Query			
Result	Score	Match Length	Description
No.	DB	ID	

1	5301	68.6	1114	16	US-10-408-765A-2119	Sequence 2119, Ap
2	2452.5	31.7	1133	14	US-10-309-851-14	Sequence 14, Appl
3	2444.5	31.6	1133	14	US-10-309-851-16	Sequence 16, Appl
4	2444.5	31.6	1135	14	US-10-309-851-12	Sequence 12, Appl
5	2406.5	31.1	512	15	US-10-108-260A-4080	Sequence 4080, Ap
6	1828	23.6	764	14	US-10-309-851-10	Sequence 10, Appl
7	1814	23.5	824	15	US-10-108-260A-3569	Sequence 3569, Ap
8	1675	21.7	785	14	US-10-309-851-24	Sequence 24, Appl
9	1645.5	21.3	752	14	US-10-171-311-48	Sequence 48, Appl
10	780.5	10.1	439	9	US-09-925-302-495	Sequence 495, App
11	780.5	10.1	439	12	US-09-925-302-495	Sequence 495, App
12	769.5	10.0	350	14	US-10-309-851-8	Sequence 8, Appli
13	677	8.8	265	14	US-10-309-851-6	Sequence 6, Appli
14	618.5	8.0	384	14	US-10-309-851-38	Sequence 38, Appli
15	618.5	8.0	386	14	US-10-309-851-22	Sequence 22, Appli
16	602	7.8	274	15	US-10-104-047-3353	Sequence 3353, Ap
17	526.5	6.8	240	14	US-10-309-851-36	Sequence 36, Appli
18	518.5	6.7	240	14	US-10-309-851-4	Sequence 4, Appli
19	469.5	6.1	1961	15	US-10-028-248A-103	Sequence 103, App
20	469.5	6.1	1961	15	US-10-107-782-103	Sequence 103, App
21	469	6.1	1960	15	US-10-236-031B-62	Sequence 62, Appli
22	469	6.1	1960	15	US-10-028-248A-104	Sequence 104, App
23	469	6.1	1960	15	US-10-107-782-104	Sequence 104, App
24	468.5	6.1	1959	15	US-10-028-248A-36	Sequence 36, Appli
25	468.5	6.1	1959	15	US-10-107-782-36	Sequence 36, Appli
26	464.5	6.0	1790	15	US-10-369-493-1586	Sequence 1586, Ap
27	463.5	6.0	1978	15	US-10-094-466-64	Sequence 64, Appli
28	463.5	6.0	2020	15	US-10-369-493-5128	Sequence 5128, Ap
29	463.5	6.0	2020	15	US-10-369-493-5129	Sequence 5129, Ap
30	462.5	6.0	1164	15	US-10-369-493-6564	Sequence 6564, Ap
31	461.5	6.0	1881	14	US-10-032-585-7646	Sequence 7646, Ap
32	461.5	6.0	1999	15	US-10-028-248A-107	Sequence 107, App
33	461.5	6.0	1999	15	US-10-107-782-107	Sequence 107, App
34	458.5	5.9	1961	15	US-10-028-248A-105	Sequence 105, App
35	458.5	5.9	1961	15	US-10-107-782-105	Sequence 105, App
36	452	5.8	1938	14	US-10-171-311-164	Sequence 164, App
37	452	5.8	1945	10	US-09-927-597-2	Sequence 2, Appli
38	452	5.8	1972	14	US-10-171-311-162	Sequence 162, App
39	452	5.8	1972	15	US-10-341-434-103	Sequence 103, App
40	452	5.8	1979	10	US-09-927-597-4	Sequence 4, Appli
41	451	5.8	1905	15	US-10-259-194A-86	Sequence 86, Appli
42	447.5	5.8	1959	15	US-10-028-248A-106	Sequence 106, App
43	447.5	5.8	1959	15	US-10-107-782-106	Sequence 106, App
44	441.5	5.7	1965	15	US-10-369-493-3279	Sequence 3279, Ap
45	441	5.7	880	15	US-10-369-493-21643	Sequence 21643, A

ALIGNMENTS

RESULT 1
US-10-408-765A-2119
; Sequence 2119, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2119
; LENGTH: 1114
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2119

Alignment Scores:				
Pred. No.:	2.63e-311	Length:	1114	
Score:	5301.00	Matches:	1054	
Percent Similarity:	97.13%	Conservative:	28	
Best Local Similarity:	94.61%	Mismatches:	32	
Query Match:	68.57%	Indels:	0	
DB:	16	Gaps:	0	
US-10-788-793-1 (1-4364) x US-10-408-765A-2119 (1-1114)				
QY	369	ATGCTGAGGACAGAGAAAACCAAGCCCGAGGTTCTGGAGGCACACTATGGATCTGCAGAA	428	
Db	1	MetLeuLysThrGluLysThrLysProGluValLeuGluAlaHisTyrGlySerAlaGlu	20	
QY	429	CCTGAGAAAGTGCTTCGGGTCTCGACCGAGATGCCATCCTTGCTCAAGAGAAGTCCATA	488	
Db	21	ProGluLysValLeuArgValLeuHisArgAspAlaIleLeuAlaGlnGluLysSerIle	40	
QY	489	GGAGAAGACGTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGMAAAGCAGAAG	548	
Db	41	GlyGluAspValTyrGluLysProIleSerGluLeuAspArgLeuGluLysGlnLys	60	
QY	549	GAGACGTACCGCCGCATGCTAGACGAGCTGCTGGCTGGAGAAAGTGTCAAGCGGCACC	608	
Db	61	GluThrTyrArgArgMetLeuGluGlnLeuLeuLeuAlaGluLysCysHisArgArgThr	80	
QY	609	GTGTACGAGCTGGAGAACGAGAAGCACACACACTGACTACATGAACAAGAGCGACGAC	668	
Db	81	ValTyrGluLeuGluAsnGluLysHisLysHisThrAspTyrMetAsnLysSerAspAsp	100	
QY	669	TTCACCAACCTGCTGGAGCGAGAGCGAGAGGTTGAAAAGCTCCTTGAACAAGAAAAA	728	
Db	101	PheThrAsnLeuLeuGluGlnGluArgGluArgLeuLysLysLeuLeuGlnGluLys	120	
QY	729	GCTTACCAGCCCGCAAAGRAAAGGAAAAACCGCTAAGCGGTCAACAACTTCGAGATGAG	788	
Db	121	AlaTyrGlnAlaArgLysGluLysGluAsnAlaLysArgLeuAsnLysLeuArgAspGlu	140	
QY	789	CTTGTGAAGCTCAAGTCTTCGCCCTCATGTTGGTGGACGAGGCGCAGATGCACATCGAG	848	
Db	141	LeuValLysLeuLysSerPheAlaLeuMetLeuValAspGluArgGlnMetHisIleGlu	160	
QY	849	CAACTGGGCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAAGCTCAGAGGAGGAGAA	908	
Db	161	GlnLeuGlyLeuGlnSerGlnLysValGlnAspLeuThrGlnLysLeuArgGluGluGlu	180	
QY	909	GA AAAACTCAAAGCGGTCACTTACAAATCCAAAGGAAGACCGCCAGAAAGCTGCTCAAGTTA	968	
Db	181	GluLysLeuLysAlaIleThrSerLysSerLysGluAspArgGlnLysLeuLysLeu	200	
QY	969	GAAGTGGACTCGACATCAGAAAGTCCAGGACCTCACTCAGAAAGCTCAGAGATGAACGCC	1028	
Db	201	GluValAspPheGluHisLysAlaSerArgPheSerGlnGluHisGluGluMetAsnAla	220	
QY	1029	AAATTGCGGAATCAAGAAATCTCAAAACCGGCAACTTCGACTCAAACTGGTTGGCTTATCG	1088	
Db	221	LysLeuAlaAsnGlnGluSerHisAsnArgGlnLeuArgLeuLysLeuValGlyLeuThr	240	
QY	1089	CAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAAGCCTTTCAGAAAGGCAGAGGAAGCTC	1148	
Db	241	GlnArgIleGluGluLeuGluGluThrAsnLysAsnLeuGlnLysAlaGluGluGlu	260	
QY	1149	CAGGAGCTGAGAGAGAAAATTGCCAAAGGGGAATGTGGAAACTCCAGTCTCATGGCGGAA	1208	
Db	261	GlnGluLeuArgAspLysIleAlaLysGlyGluCysGlyAsnSerSerLeuMetAlaGlu	280	
QY	1209	GTGGAGAGTCTGCGCAAGCGCGTGTGATGAGATGGAGGCAAGGATGAAGAGATCACGAAG	1268	
Db	281	ValGluAsnLeuArgLysArgValLeuGluMetGluGlyLysAspGluGluIleThrLys	300	
QY	1269	ACCGAGCCCGAGTGC CGGAGCTGAAGAAAGAGCTCCAAGAGGAAGAACCCACAGCAAG	1328	

Db	301	ThrGluSerGlnCysArgGluLeuArgLysLysLeuGlnGluGluGluHisSerLys	320	
QY	1329	GAAC TTAGACTAGAAGCTGAGAAAGCTGCAGAAAGAGGATGTCTGAGCTGGAGAAGCTGGAG	1388	
Db	321	GluLeuArgLeuGluValGluLysLeuGlnLysArgMetSerGluLeuGluLysLeuGlu	340	
QY	1389	GAAGCGTTTCAGCCGGAGTAAGTCGGAATGTCACCCAGCTCCATCTGAACCTGGAGAAGGAG	1448	
Db	341	GluAlaPheSerLysSerLysSerGluCysThrGlnLeuHisLeuAsnLeuGluLysGlu	360	
QY	1449	AAGAACCTAACCAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCCAGTTAAAGAA	1508	
Db	361	LysAsnLeuThrLysAspLeuLeuAsnGluLeuGluValValLysSerArgValLysGlu	380	
QY	1509	CTCGAATGCTCCGAGAGTAGACTGGAGAAGGCCGAGTTAAGCCTCAAGAGATGACCTTACA	1568	
Db	381	LeuGluCysSerGluSerArgLeuGluLysAlaGluLeuSerLeuLysAspAspLeuThr	400	
QY	1569	AAGCTGAAGTCTTCACTGTGATGCTGGTGGATGAGAGGAAAAATATATGATGGAGAAAAATA	1628	
Db	401	LysLeuLysSerPheThrValMetLeuValAspGluArgLysAsnMetMetGluLysIle	420	
QY	1629	AAGCAAGAAGAGAGAGAAAGTGGATGGGTTGAATAAAAACTTTAAGGTGGAGCAGGGAATA	1688	
Db	421	LysGlnGluGluArgLysValAspGlyLeuAsnLysAsnPheLysValGluGlnGlyLys	440	
QY	1689	GT CATGGATGTACGGAAAAAGCTTAATCGAGGAAAGCAAGAAAGCTTTTAAAAACTCAAATCT	1748	
Db	441	ValMetAspValThrGluLysLeuIleGluGluSerLysLysLeuLeuLysLeuLysSer	460	
QY	1749	GAATGGAGGAAAAGGAGTACAGTCTGACAAAGGAGAGGGATGAGTGTGGTAAACTG	1808	
Db	461	GluMetGluGluLysValTyrAsnLeuThrArgGluArgAspGluLeuIleGlyLysLeu	480	
QY	1809	AGGAGCGAAGAAAGAGTCCCTGTGAAC TGAGCTGCAGTGTAGACTTACTATAAGAACCGG	1868	
Db	481	LysSerGluGluGluLysSerSerGluLeuSerCysSerValAspLeuLeuLysLysArg	500	
QY	1869	CTTGATGGCATAGAGGAGGTAGAAAGGGAATAAACCGAGGTAGGTGTCGAAGGGGTCT	1928	
Db	501	LeuaspGlyIleGluGluValGluArgGluIleThrArgGlyArgSerArgLysGlySer	520	
QY	1929	GAGTTCACCTGCCCGGAAGACAATAAGATCAGAGNACTAACCGTTGAAATCGAGAGACTG	1988	
Db	521	GluLeuThrCysProGluAspAsnLysIleLysGluLeuThrLeuGluIleGluArgLeu	540	
QY	1989	AAGAAACGGCTCCACGAGTTGGAGTGGTGGAGGGGACTTGTATGAAGACCCGAGGACGAA	2048	
Db	541	LysLysArgLeuGlnGlnLeuGluValValGluGlyAspLeuMetLysThrGluAspGlu	560	
QY	2049	TATCACCAGTTGGAGCAGAAGTTCAGAAACCGAGCAGGATTAAGGCAAACTTCCTCTCCAG	2108	
Db	561	TyrAspGlnLeuGluGlnLysPheArgThrGluGlnaspLysAlaasnPheLeuSerGln	580	
QY	2109	CAGCTCGAGGAATCAAAACACCACCAATGGCCCAAGCACAAGCCATAGAGAAAGGGAGGCC	2168	
Db	581	GlnLeuGluGluIleLysHisGlnIleAlaLysAsnLysAlaIleGluLysGlyGluVal	600	
QY	2169	GTGAGCCAGGAAGCCGAAC TCGACACACAGGTTTTCGGCTGGAGGAGGCTAAAAAGTCGTGAT	2228	
Db	601	ValSerGlnGluAlaGluLeuArgHisArgPheArgLeuGluGluAlaLysSerArgAsp	620	
QY	2229	TTACAGGCCGAGTGCAGGCTCTCAAGGAGAAGATCCACGAGCTGATGAACAAGAAAGAC	2288	
Db	621	LeuLysAlaGluValGlnAlaLeuLysGluLysIleHisGluLeuMetAsnLysGluAsp	640	
QY	2289	CAGCTGTCTCAGCTCCAAGTCGACTATTTCGGTCTCTCAGCAAAAGATTTATGGAAGAAAGAA	2348	
Db	641	GlnLeuSerGlnLeuGlnValAspTyrSerValLeuGlnGlnArgPheMetGluGluGlu	660	
QY	2349	ACTAAGAACAAAGACATGGGGAGGGAGGTCCTCAATCTGACCAAGGAGCTAGAGCTTTCC	2408	
Db	661	AsnLysAsnLysAsnMetGlyGlnGluValLeuAsnLeuThrLysGluLeuGluLeuSer	680	

QY 2409 AAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAAGGATGGTGGAGCTGCCT 2468
Db |||||||
681 LysArgTyrSerArgAlaLeuArgProSerValAsnGlyArgArgMetValAspValPro 700
QY 2469 GTGGCCTCCACTGGGTGCAGACCAGCGCGGTGTGCGGGGATGCTGCGGAGGAGGAGACC 2528
Db ||| |||||||
701 ValThrSerThrGlyValGlnThrAspAlaValSerGlyGluAlaAlaGluGluGluThr 720
QY 2529 CCGGCTGTGTTTCATTCGCAAAATCCTTCCAGGAGGAGAAATCACATCATGAGTAATCTTGA 2588
Db |||||||
721 ProAlaValPheIleArgLysSerPheGlnGluAsnHisIleMetSerAsnLeuArg 740
QY 2589 CAGGTAGGCCTGAAGAAACCCATGGAACGGTCTCGGTCTCTCGACAGGTATCCCCCAGCA 2648
Db |||||||
741 GlnValGlyLeuLysLysProValGluArgSerValLeuAspArgTyrProProAla 760
QY 2649 GCGAATGAGCTCACCATGAGGAAGTCTTGGATTCCTTGGATGAGAAAAAGAGAAAAACGGT 2708
Db |||||||
761 AlaAsnGluLeuThrMetArgLysSerTrpIleProTrpMetArgLysArgGluAsnGly 780
QY 2709 CCTTCCACTCCGAGGAGAAAGGGCCCGCCAGGCCAAACAGGGTGCAGGGCACCCCGGGAG 2768
Db |||||||
781 ProSerIleThrGlnGluLysGlyProArgThrAsnSerSerProGlyHisProGlyGlu 800
QY 2769 CTGGTCCTAGCACCAAGCAGGGCCAGCCCCCTACACATCCGTGTGACACCACGATCATGAG 2828
Db ::::: |||||||
801 ValValLeuSerProLysGlnGlyGlnProLeuHisIleArgValThrProAspHisGlu 820
QY 2829 AACAGCACTGCCACCTGGAGATCACAAAGCCCAATCTGAAGAGTTTTTCTCTAGTACC 2888
Db |||||||
821 AsnSerThrAlaThrLeuGluIleThrSerProThrSerGluGluPhePheSerSerThr 840
QY 2889 ACCGTCAATTCTACCTTAGGCAACACAGAAACCAAGAATAACCATATTATCCATCACCCAAT 2948
Db |||||||
841 ThrValIleProThrLeuGlyAsnGlnLysProArgIleThrIleIleProSerProAsn 860
QY 2949 GTCATGTCGAAAAAGCCCAAAAGTGCAGATCCTACTCTCGGCCCAGAACGAGCATGTCC 3008
Db |||||||
861 ValMetProGlnLysGlnLysSerGlyAspThrThrLeuGlyProGluArgAlaMetSer 880
QY 3009 CCTGTCACGATTACTACTATTTCAGAGAGAAGAGCCCGGAAGGTGGAAGGAGCGCCTTT 3068
Db |||||||
881 ProValThrIleThrThrPheSerArgGluLysThrProGluSerGlyArgGlyAlaPhe 900
QY 3069 GCCGACAGGCCTGCATCCCCCATCCAAATCATAGCGGTGTCAACATCTGCAGCTCCCACT 3128
Db |||||||
901 AlaAspArgProThrSerProIleGlnIleMetThrValSerThrSerAlaAlaProAla 920
QY 3129 GAAATCGTGTCTCTCTCGTAATCTCAGGAAGTGCCATATGGGAAGGACTATCCTCAAAGTC 3188
Db |||||||
921 GluIleAlaValSerProGluSerGlnGluMetProMetGlyArgThrIleLeuLysVal 940
QY 3189 ACCCCGGAACAAACTGTTCCAGCCCCCGTGCGAAGTACAACCTCCAATGCTAATATC 3248
Db |||||||
941 ThrProGluLysGlnThrValProThrProValArgLysTyrAsnSerAsnAlaAsnIle 960
QY 3249 ATCACCACGGAAGACAATAAAATTACATTCACCTGGGTTCTCAGTTTAAAGCGATCTCCT 3308
Db |||||||
961 IleThrThrGluAspAsnLysIleHisIleHisLeuGlySerGlnPheLysArgSerPro 980
QY 3309 GGGCCTGCCGCTGAAGCGGTGAGCCCCAGTTATACCGTCCGGCCTGTCAACGTGACAGCG 3368
Db ||| ::| |||||||
981 GlyThrSerGlyGluGlyValSerProValIleThrValArgProValAsnValThrAla 1000
QY 3369 GAGAAGGAGGTTTCTACAGGCACAGTCTCTTCGTTCTCCAGGAACCACTCTCTTCAAGA 3428
Db |||||||
1001 GluLysGluValSerThrGlyThrValLeuArgSerProArgAsnHisLeuSerSerArg 1020
QY 3429 CCCGGTGTAGCAAAAGTGACCAGCACTATAACTATAACCCCGGTACAAACGTCATCCACA 3488
Db |||||||
1021 ProGlyAlaSerLysValThrSerThrIleThrIleThrProValThrThrSerSerAla 1040

QY 3489 CGAGGAACCCCAATCAGTGTCCAGGACAAAGATGGGTATCTCAGCGGCCTACCCCCACCCGC 3548
Db |||||||
1041 ArgGlyThrGlnSerValSerGlyGlnAspGlySerSerGlnArgProThrProThrArg 1060
QY 3549 ATTCTATGTCAAAAGGTATGAAAGCTGGAAGCCAGTAGTGGCAGCCTCAGGAGCAGGA 3608
Db |||||||
1061 IleProMetSerLysGlyMetLysAlaGlyLysProValValAlaAlaProGlyAlaGly 1080
QY 3609 AATCTGACCAAAATTCAGCCTCGAGCTCGAGCTGAGACTCAGTCTATGAAAAATAGAGTGAAGAAA 3668
Db |||||||
1081 AsnLeuThrLysPheGluProArgAlaGluThrGlnSerMetLysIleGluLeuLysLys 1100
QY 3669 TCTGCAGCCAGCAGCACTGCCTCTCTTGGAGGGGGAAGGGC 3710
Db |||||||
1101 SerAlaAlaSerSerThrThrSerLeuGlyGlyGlyLysGly 1114
RESULT 2
US-10-309-851-14
; Sequence 14, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
; CURRENT APPLICATION NUMBER: US/10/309,851
; CURRENT FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1133
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-851-14
Alignment Scores:
Pred. No.: 1.93e-139 Length: 1133
Score: 2452.50 Matches: 542
Percent Similarity: 64.40% Conservative: 214
Best Local Similarity: 46.17% Mismatches: 337
Query Match: 31.72% Indels: 81
DB: 14- Gaps: 19
US-10-788-793-1 (1-4364) x US-10-309-851-14 (1-1133)
QY 75 ATGAGATCACGAAATCAAGGTGGAGAAAGTTTCATCTAACGGGCATGTCTCCTGCCCAAG 134
Db |||||||
1 MetArgSerArg-----GlySerAspThrGluGlySerAlaGlnLysLysPheProArg 18
QY 135 TCCTCCATCATCAGCAGTGATGGTGGTAAGGCCCTCAGAAGATGCAAAAAAGAACAAAG 194
Db ::| |||||||
19 HisThr-----LysGlyHisSerPheGlnGlyProLysAsnMet 31
QY 195 GCCAATCGGAAGGAGGAG---GATGTATGGCTTCCGGAACCTATCAAAAGGCACCTCAAA 251
Db ::||| ::|||
32 LysHisArgGlnAspLysAspSerProSerGluSerAspVal-----IleLeu 48
QY 252 CCATCTGGAGAAAGTGAGAAA-----AAGACTAAGAAAGTCTGTGAGTTATCC 299
Db ||| ::|||
49 ProCysProLysAlaGluLysProHisSerGlyAsnGlyHisGlnAlaGluAspLeuSer 68
QY 300 AAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGAGTTGCAGGTCGAGAAGAT 359
Db ::||| ::|||
69 ArgAspAspLeuLeuPheLeuLeuSerIleLeuGluGlyGluLeuGlnAlaArgAspGlu 88
QY 360 GTCATCCACATGCTGAGGACAGAGAAAAACCAAGCCCGAGGTTCTGGAGGCACACTATGA 419
Db |||||||
89 ValIleGlyIleLeuLysAlaGluLysMetAspLeuAlaLeuLeuGluAlaGlnTyrGly 108
QY 420 TCTGCAGAACCTCAGAAAAGTGTTCGGGTCTCGCACCGAGATGCCATCCTTGCTCAAGAG 479
Db |||||
109 PheValThrProLysLysValLeuGluAlaLeuGlnArgAspAlaPheGlnAlaLysSer 128

QY	480	AAGTCCATAGGAGAAAGACGTCTATGAGAAAACCTATCTCAGAGCTGGACAGACTGGAGGAA	539
Db	129	ThrProTrpGlnGluAspIleTyrGluLysProMetAsnGlnLeuAspLysValValGlu	148
QY	540	AAGCAGAAAGGAGACGTACCCGCGCATGCTAGACAGCTGCTGCTGGCTGAGAGTGTCA	599
Db	149	LysHisLysGluSerTyrArgArgIleLeuGlyGlnLeuLeuValAlaGluLysSerHis	168
QY	600	AGGCGCACCGTGTACGAGCTGGAGAAACGAGAACACACACTGACTACATGAACAAG	659
Db	169	ArgGlnThrIleLeuGluLeuGluGluLysArgLysHisLysGluTyrMetGluLys	188
QY	660	AGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGGAGAGGTTGAAAAAGCTCCTTGAA	719
Db	189	SerAspGluPheIleCysLeuLeuGluGlnGluCysGluArgLeuLysLysLeuIleAsp	208
QY	720	CAAGAAAAAGCTTACCAGCCCGCAAGAAAGAAAACGCTAAAGCGGCTCAACAAACTT	779
Db	209	GlnGluIleLysSerGlnGluGluLysGluGlnGluLysGluLysArgValThrThrLeu	228
QY	780	CGAGATGAGCTTGTAAGCTCAAGTCTTCGCCCTCATGTTGGTGGACGAGGCGAGATG	839
Db	229	LysGluGluLeuThrLysLeuLysSerPheAlaLeuMetValValAspGluGlnArg	248
QY	840	CACATCGAGCAACTGGCCCTCGAGATCAGAAAGTCCAGGACCTCACTCAGAAAGCTGAG	899
Db	249	LeuThrAlaGlnLeuThrLeuGlnArgGlnLysIleGlnGluLeuThrThrAsnAlaLys	268
QY	900	GAGGAGGAAGAAAAACTCAAAAGCGGTCACTTACAAATCCAAAGGAAGACCGCCAGAAGCTG	959
Db	269	GluThrHisThrLysLeuAlaLeuAlaGluAlaArgValGlnGluGlnLysAla	288
QY	960	CTCAAGTTAGAAGTGGACTTCGAACACACAGGCCCTCGAGGTTTCCCAGGACGACGAAG	1019
Db	289	ThrArgLeuGluLysGluLeuGlnThrGlnThrThrLysPheHisGlnAspGlnAspThr	308
QY	1020	ATGAACGCCAAATTTGGCGAATCAAGAATCTCACAAACCGGCAACTTCGACTCAAACTGGTT	1079
Db	309	IleMetAlaLysLeuThrAsnGluAspSerGlnAsnArgGlnLeuGlnLysLeuAla	328
QY	1080	GGCTTATCGCAAGGATTGAGGAGCTGGAAGAGACCAATPAAAGCCTTCAGAAAGGCAGAG	1139
Db	329	AlaLeuSerArgGlnIleAspGluLeuGluGluThrAsnArgSerLeuArgLysAlaGlu	348
QY	1140	GAAGAGCTCCAGGAGCTGAGAGAGAAAATTGCCAAAGGGGAATGTGGAAACTCCAGTCTC	1199
Db	349	GluGluLeuGlnAspIleLysGluLysIleSerLysGlyGluTyrGlyAsnAlaGlyIle	368
QY	1200	ATGGCGGAAGTGGAGAGTCTCGCAAGCGCGTCTTGAGATGGAGGCGAAGGATGAAGAG	1259
Db	369	MetAlaGluValGluGluLeuArgLysArgValLeuAspMetGluGlyLysAspGluGlu	388
QY	1260	ATCACGAAACCGAGGCCAGTCCCGGAGCTGAAGAAGAGCTCCAAAGAGGAAGAACAC	1319
Db	389	LeuIleLysMetGluGluGlnCysArgAspLeuAsnLysArgLeuGluArgGluThrLeu	408
QY	1320	CACAGCAAGAACTTAGACTAGAAAGTGGAGAAGCTGCAGAAGAGGATGCTCTGAGCTGGAG	1379
Db	409	GlnSerLysAspPheLysLeuGluValGluLysLeuSerLysArgIleMetAlaLeuGlu	428
QY	1380	AAGCTGGAGGAAGCTTCAGCCGGAGTAAGTCGGAATGCCACCCAGCTCCATCTGAACCTG	1439
Db	429	LysLeuGluAspAlaPheAsnLysSerLysGlnGluCysTyrSerLeuLysCysAsnLeu	448
QY	1440	GAGAAGGAGAAACCTTAACCAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGA	1499
Db	449	GluLysGluArgMetThrThrLysGlnLeuSerGlnGluLeuGluSerLeuLysValArg	468
QY	1500	GTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAGGCCGAGTTAAGCCCTCAAGAT	1559
Db	469	IleLysGluLeuGluAlaIleGluSerArgLeuGluLysThrGluPheThrLeuLysGlu	488

QY	1560	GACCTTACAAAGCTGAAGTCCTTCACCTGTGATGCTGGTGGATGAGAGGAAAAATATGATG	1619
Db	489	AspLeuThrLysLeuLysThrLeuThrValMetPheValAspGluArgLysThrMetSer	508
QY	1620	GAGAAAAATAAAGCAAGAGAGAGGAAAGTGGATGGGTTGAATAAAAACTTTAAGGTGGAG	1679
Db	509	GluLysLeuLysLysThrGluAspLysLeuGlnAlaIaSerSerGlnLeuGlnValGlu	528
QY	1680	CAGGGAAGAGTCATGGATGTGACGGAAAAGCTAATCGAGGAAAAAGCAAGAGCTTTTAAA	1739
Db	529	GlnAsnLysValThrThrValThrGluLysLeuIleGluGluThrLysArgAlaLeuLys	548
QY	1740	CTCAAACTGAAATGGAGGAAAAAGGAGTACAGTCTGACAAAAGGAGGAGGATGAGCTGATG	1799
Db	549	SerLysThrAspValGluGluLysMetTyrSerValThrLysGluArgAspAspLeuLys	568
QY	1800	GGTAACTGAGGAGCGGAAGAAAGTCTCTGTGAACCTGAGCTGCAGTCTAGACTTACTA	1859
Db	569	AsnLysLeuLysAlaGluGluGluLysGlyAsnAspLeuLeuSerArgValAsnMetLeu	588
QY	1860	AAGAAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGGGAATAAACCGAGGTAGG-----	1913
Db	589	LysAsnArgLeuGlnSerLeuGluAlaIleGluLysAspPheLeuLysAsnLysLeuAsn	608
QY	1914	-----TCGTGCAAGGGGTCTGAGTTCACTTCACCTGCCCGGAAGACAAATAAGATCAGAGAACTA	1967
Db	609	GlnAspSerGlyLysSerThrThrAlaLeuHisGlnGluAsnAsnLysIleLysGluLeu	628
QY	1968	ACGCTTGAATCGAGAGACTGAAGAAACGGCTCCAGAGTTGGAGGTGGTGGAGGGGGAC	2027
Db	629	SerGlnGluValGluArgLeuLysLeuLysLeuLysAspMetLysAlaIleGluAspAsp	648
QY	2028	TTGATGAAGACCGAGGACGAATATGACCAGTTTGGAGCAGAAAGTTTCAGAACCGAGCAGGAT	2087
Db	649	LeuMetLysThrGluAspGluTyrGluThrLeuGluArgArgTyrAlaAsnGluArgAsp	668
QY	2088	AAGGCAAAACCTTCTCTCCAGCAGCTCGAGGAAATCAAATGGCCAAAGCACAAAA	2147
Db	669	LysAlaGlnPheLeuSerLysGluLeuGluHisValLysMetGluLeuAlaLysTyrLys	688
QY	2148	GCCATAGAGAAAGGGGAGGCCGTGAGCCAGGAAGCCGAACTGCGACACAGGTTTCGGCTG	2207
Db	689	LeuAlaGluLysThrGlu--ThrSerHisGluGlnTrpLeuPheLysArgLeuGlnGlu	707
QY	2208	GAGGAGGCTAAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAGGAGAGATCCAC	2267
Db	708	GluGluAlaLysSerGlyHisLeuSerArgGluValAspAlaLeuLysGluLysIleHis	727
QY	2268	GAGCTGATGAACAAGGAAGACCAAGTGTCTCAGTCCAAAGTCGACTATTTCGTCCTTCAG	2327
Db	728	GluTyrMetAlaThrGluAspLeuIleCysHisLeuGlnGlyAspHisSerValLeuGln	747
QY	2328	CAAGATTTATGGAAGAGAAACCTAAGAACCAAGAACATGGGAGGAGGTCCTCAATCTG	2387
Db	748	LysLysLeuAsnGlnGlnGluAsnArgAsnArgAspLeuGlyArgGluIleGluAsnLeu	767
QY	2388	ACCAAGGAGCTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGCGCGAGTGGGAACGCG	2447
Db	768	ThrLysGluLeuGluArgTyrArgHisPheSerLysSerLeuArgProSerLeuAsnGly	787
QY	2448	CGAAGGATGGTGACGTGCCTGTGGCCTCCACTGGGGTGCAGCCGAGCGGTGTCGGG	2507
Db	788	ArgArgIleSerAspProGlnValPheSerLysGluValGlnThrGluAlaVal-----	805
QY	2508	GATGCTCGGAGGAGGAGACCCCG-----GCTGTG	2537
Db	806	-----AspAsnGluProProAspTyrLysSerLeuIleProLeuGluArgAlaVal	822
QY	2538	TTCATTCGCAATCCTTCCAGGAG---GAAAAATCACATCATAGTAATCTTCGACAGGTA	2594
Db	823	IleAsnGlyGlnLeuTyrGluGluSerGluAsnGlnAsp-----	835
QY	2595	GGCCTGAAGAAACCCCATGGAACGGTCTCTCGGTCTCTC-----GACAGGTAT	2639

```

Db      836  ----GluAspProAsnAspGluGlySerValLeuSerPheLysCysSerGlnSerThr 853
QY      2640  CCCCAGCAGCGAATGAGCTCACCATTGAGGAAGTCTTGGATTCTCTGGATGAGAAAAAGA 2699
Db      854  ProCysProValAsn-----ArgLysLeuTrpIleProTrpMetLysSerLys 869
QY      2700  GAA-----AACGGTCCTTCCACTCCGCGCAGGAGAAAGGGCCCGCCAGCCCAACCCAG 2747
Db      870  GluGlyHisLeuGlnAsnGly-----LysMetGlnThrLysProAsnAla 884
QY      2748  GGTGCAGGGCACCCCGGGAGCTGGTCCTAGCACCAAGAGCGGCCAGCCCTACACATC 2807
Db      885  AsnPheValGlnProGlyAspLeuValLeuSerHisThrProGlyGlnProLeuHisIle 904
QY      2808  CGTGTGACACCAGATATGAGAAACAGCACTGCCACCTGGAGATCACAAAGCCCACTCT 2867
Db      905  LysValThrProAspHisValGlnAsnThrAlaThrLeuGluIleThrSerProThrThr 924
QY      2868  GAA-----GAGTTTCTCTAGTACCACCGTCATTCTCTACCTTAGGCAACCAGAAACCA 2921
Db      925  GluSerProHisSerTyThrSerThrAlaValIleProAsnCysGlyThrProLysGln 944
QY      2922  AGAATAACCATATTCCATCACCCCAATGTCTGTCGCAAAAGCCCAAA---AGTGCAGAT 2978
Db      945  ArgIleThrIleLeuGlnAsnAlaSerIleThrProValLysSerLysThrSerThrGlu 964
QY      2979  CCTACTCTCGGCCCCAGAACGAGCCATGTCCCCCTGTACAGATTACTACTATTTCAGAGAG 3038
Db      965  AspLeuMetAsnLeuGluGlnGlyMetSerProIleThrMetAlaThrPheAlaArgAla 984
QY      3039  AAGAGCCCGGAAGGTGGAAGGAGCGCCTTTGCCGACAGGCGCTGCATCCCCCATCCAAATC 3098
Db      985  GlnThrProGluSerCysGlySerLeuThrProGluArgThrMetSerProIleGlnVal 1004
QY      3099  ATGACGGTGTCAACATCTGCAGCTCCCACCTGAAATCGTGTCTCTCTGAAATCTCAGGAA 3158
Db      1005  LeuAlaValThrGlySerAlaSerSerProGluGlnGlyArgSerProGluProThrGlu 1024
QY      3159  GTGCCATATGGGAAGGACTATCCTCAAAGTCAACCCCGGAAAAACAACTGTTCCAGCCCCC 3218
Db      1025  IleSerAlaLysHisAlaIlePheArgValSerProAspArgGlnSerSerTrpGlnPhe 1044
QY      3219  GTGCGGAAGTACAACCTCCAATGCTAATATCATCATCACCGGAAGACAAATAAATTCACATT 3278
Db      1045  GlnArgSerAsnSerAsnSerSerValIleThrThrGluAspAsnLysIleHisIle 1064
QY      3279  CACCTGGGTTCTCAGTTTAAGCGATCTCCTGGGCTGCCGCTGAAGGCGTGAGCCCGAGTT 3338
Db      1065  HisLeuGlySerProTyThrMetGlnAla-----ValAlaSerProValArgProAla 1081
QY      3339  ATCACCGTCCGGCTGTCAACGTGACAGCGGAGAGGAGGTTTCTACAGGCACAGTCCTT 3398
Db      1082  SerProSerAlaProLeuGlnAspAsnArgThrGlnGlyLeuIleAsnGlyAlaLeuAsn 1101
QY      3399  CGCTCTCCCAGGAACCACTCTCTTCAAGACCCCGGTGCTAGCAAAGTGACCCAGCACTATA 3458
Db      1102  LysThr-----ThrAsnLysValThrSerSerIle 1111
QY      3459  ACTATAACCCCGGTCAACAGTCATCCACACGAGGAACCCAA 3500
Db      1112  ThrIleThrProThrAlaThrProLeuProArgGlnSerGln 1125

RESULT 3
US-10-309-851-16
; Sequence 16, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
```

```

; CURRENT APPLICATION NUMBER: US/10/309,851
; CURRENT FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 1133
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-851-16

Alignment Scores:
Pred. No.:      5.87e-139      Length:      1133
Score:          2444.50        Matches:    541
Percent Similarity: 64.31%      Conservative: 214
Best Local Similarity: 46.08%    Mismatches: 338
Query Match:    31.62%         Indels:     81
DB:             14             Gaps:       19

US-10-788-793-1 (1-4364) x US-10-309-851-16 (1-1133)

QY      75  ATGAGATCACGAATCAAGGTGGAGAAAGTTTCATCTAACGGGCATGTCTCTGCCCCCAAG 134
Db      1  MetArgSerArg-----GlySerAspThrGluGlySerAlaGlnLysLysPheProArg 18
QY      135  TCCTCCATCATCAGCAGTGATGGTGGTAAGGCCCCCTCAGAAGATGCAAAAAAGAACAAAG 194
Db      19  HisThr-----LysGlyHisSerPheGlnGlyProLysAsnMet 31
QY      195  GCCAATCGGAAGGAGGAG--GATGTCTATGGCTTCCGGAACCTATCAAAAGGCACCTCAAA 251
Db      32  LysHisArgGlnGlnAspLysAspSerProSerGluSerAspVal-----IleLeu 48
QY      252  CCATCTGGAGAAAGTGAGAAA-----AAGACTAAGAAAGTCTGTGGAGTTATCC 299
Db      49  ProCysProLysAlaGluLysProHisSerGlyAsnGlyHisGlnAlaGluAspLeuSer 68
QY      300  AAGGAGGACCTCATCCAGCTCCTGAGTATCATTGAAGGGGAGTTGCAGGCTCGAGAAAGAT 359
Db      69  ArgAspAspLeuLeuPheLeuLeuSerIleLeuGluGlyLeuGlnAlaArgAspGlu 88
QY      360  GTCATCCACATGCTGAGGACAGAGAAAAACCAAGCCCGAGGTTCTTGAGGACACACTATGGA 419
Db      89  ValIleGlyIleLeuLysAlaGluLysMetAspLeuAlaLeuLeuGluAlaGlnTyGly 108
QY      420  TCTGCAGAACCTGAGAAAGTGTTCGGGTCTCTGCACCGAGATGCCATCCTTGCTCAAGAG 479
Db      109  PheValThrProLysLysValLeuGluAlaLeuGlnArgAspAlaPheGlnAlaLysSer 128
QY      480  AAGTCCATAGGAGAAAGACGTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAA 539
Db      129  ThrProTrpGlnGluAspIleTyThrGluLysProMetAsnGluLeuAspLysValValGlu 148
QY      540  AAGCAGAAAGGAGACGTACCGCCGATGCTAGACGAGCTGCTGCTGGCTGAGAGTGTCTAC 599
Db      149  LysHisLysGluSerTyThrArgArgIleLeuGlyGlnLeuLeuValAlaGluLysSerArg 168
QY      600  AGCGCACCGTGTACGAGCTGGAGAACGAGAGACCAAGCACACTGACTACATGAACAAG 659
Db      169  ArgGlnThrIleLeuGluLeuGluGluGluGlnGluCysGluArgLeuLysLysLeuIleAsp 188
QY      660  AGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGAGAGGTTTGAATAAGCTCTTGAA 719
Db      189  SerAspGluPheIleCysLeuLeuGluGlnGluCysGluArgLeuLysLysLeuIleAsp 208
QY      720  CAAGAAAAAGCTTACCAAGCCCGCAAGAAAAAGAAAAACGCTAAGCGGCTCAACAACTT 779
Db      209  GlnGluIleLysSerGlnGluLysGluGlnGluLysGluLysArgValThrThrLeu 228
QY      780  CGAGATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCATGTTGGTGGACGACGAGGCAGATG 839
Db      229  LysGluGluLeuThrLysLeuLysSerPheAlaLeuMetValValAspGluGlnGlnArg 248
QY      840  CACATCGAGCAACTGGGCGCTGCAGAGTGCAGAAAGTCCAGGACCTCACTCAGAAAGCTGAGG 899
```

Db	249	LeuThrAlaGlnLeuThrLeuGlnArgGlnLysIleGlnGluLeuThrThrAsnAlaLys	268
QY	900	GAGGAGGAAGAAAACTCAAAGCGGTCACTTACAAATCCAAGGAAGACCGCAGAAAGCTG	959
Db	269	GluThrHisThrLysLeuAlaLeuAlaGluAlaArgValGlnGluGluGlnLysAla	288
QY	960	CTCAAGTTAGAGTGGACTTCGAACACAAGGCTCGAGGTTTCCCGAGGACGACGAAGAG	1019
Db	289	ThrArgLeuGluLysGluLeuGlnThrGlnThrThrLysPheHisGlnAspGlnAspThr	308
QY	1020	ATGAACGCCAAATTGGCGAATCAAGAATCTCACAAACCGGCAACTTCGACTCAAACTGGTT	1079
Db	309	IleMetAlaLysLeuThrAsnGluAspSerGlnAsnArgGlnLeuGlnGlnLysLeuAla	328
QY	1080	GGCTTATCGCAAAAGGATTGAGGAGCTGGAAGAGACCAATAAAGCCCTTCAGAAGGCAGAG	1139
Db	329	AlaLeuSerArgGlnIleAspGluLeuGluThrAsnArgSerLeuArgLysAlaGlu	348
QY	1140	GAAGAGCTCCAAGGAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGTGAAACTCCAGTCTC	1199
Db	349	GluGluLeuGlnAspIleLysGluLysIleSerLysGlyGluTyrGlyAsnAlaGlyIle	368
QY	1200	ATGGCGGAAGTGGAGAGTCTCCGCAAGCGCGTGTGATGATGGAGGGCAAGGATGAAGAG	1259
Db	369	MetAlaGluValGluGluLeuArgLysArgValLeuAspMetGluGlyLysAspGluGlu	388
QY	1260	ATCACGAAGACCGGAGCCCGAGTCCGGGAGCTGAAGAAGAAGCTCCAAGAGGAAGAACAC	1319
Db	389	LeuIleLysMetGluGluGlnCysArgAspLeuAsnLysArgLeuGluArgGluThrLeu	408
QY	1320	CACAGCAAGAACTTAGACTAGAAAGTGGAGAAGCTGCAGAAGAGGATGCTGAGCTGGAG	1379
Db	409	GlnSerLysAspPheLysLeuGluValGluLysLeuSerLysArgIleMetAlaLeuGlu	428
QY	1380	AAGCTGGAGGAAGCGTTTCAGCCGGAGTAAAGTCGGAATGCACCCAGCTCCATCTGAACCTG	1439
Db	429	LysLeuGluAspAlaPheAsnLysSerLysGlnGluCysTyrSerLeuLysCysAsnLeu	448
QY	1440	GAGAAGGAGAAACCTTAACCAAAAGACCTGCTGAACGAGCTGGAGGTGTCGAAGAGTCGA	1499
Db	449	GluLysGluArgMetThrThrLysGlnLeuSerGlnGluLeuGluSerLeuLysValArg	468
QY	1500	GTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAGCCCGAGTTAAGCCTCAAAGAT	1559
Db	469	IleLysGluLeuGluAlaIleGluSerArgLeuGluLysThrGluPheThrLeuLysGlu	488
QY	1560	GACCTTACAAAGCTGAAGTCCTTCACTGTGATGCTGGTGGATGAGAGGAAAAATATGATG	1619
Db	489	AspLeuThrLysLeuLysThrLeuThrValMetPheValAspGluArgLysThrMetSer	508
QY	1620	GAGAAAATAAAGCAAGAGAGAGGAAGTGGATGGGTTGAAATAAAAACTTTAAGGTGGAG	1679
Db	509	GluLysLeuLysLysThrGluAspLysLeuGlnAlaIaSerSerGlnLeuGlnValGlu	528
QY	1680	CAGGGAAGAGTCATGGATGTGACGGAAAAAGCTAATCGAGGAAAGCAAGAAGCTTTTAAAA	1739
Db	529	GlnAsnLysValThrThrValThrGluLysLeuIleGluThrLysArgAlaLeuLys	548
QY	1740	CTCAAACTCTGAAATGGAGGAAAAAGGAGTACAGTCTGCACAAAGGAGAGGATGAGCTGATG	1799
Db	549	SerLysThrAspValGluLysMetTyrSerValThrLysGluArgAspAspLeuLys	568
QY	1800	GGTAACTGAGGAGCGAAGAAGAGTCCTGTGAACGTGAGCTGCAGTGTAGACTTACTA	1859
Db	569	AsnLysLeuLysAlaGluGluGluLysGlyAsnAspLeuLeuSerArgValAsnMetLeu	588
QY	1860	AAGAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGGGGAAATAAACCGAGGTAGG-----	1913
Db	589	LysAsnArgLeuGlnSerLeuGluAlaIleGluLysAspPheLeuLysAsnLysLeuAsn	608
QY	1914	-----TCGTGCAAGGGGTCTGAGTTTCCCTGCCCGGAAGACAATAAGATCAGAGAACTA	1967

Db	609	GlnAspSerGlyLysSerThrThrAlaLeuHisGlnGluAsnAsnLysIleLysGluLeu	628
QY	1968	ACGCTTGAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTCGAGGTTGGAGGGGGAC	2027
Db	629	SerGlnGluValGluArgLeuLysLeuLysLeuLysAspMetLysAlaIleGluAspAsp	648
QY	2028	TTGATGAAGACCGAGGCGAAATATGACCAGTTTGAGCGAGAAAGTTTCAGAACCGAGCAGGAT	2087
Db	649	LeuMetLysThrGluAspGluTyrGluThrLeuGluArgArgTyrAlaAsnGluArgAsp	668
QY	2088	AAGGCAAACTTCCTCTCCAGCAGCTCGAGGAAATCAAACACCACAAATGGCCAAAGCACAAA	2147
Db	669	LysAlaGlnPheLeuSerLysGluLeuGluHisValLysMetGluLeuAlaLysTyrLys	688
QY	2148	GCCATAGAGAAAAGGGAGCGCGTGAGCCAGGAAGCCGAACTGCGACACAGGTTTCGGCTG	2207
Db	689	LeuAlaGluLysThrGlu---ThrSerHisGluGlnTrpLeuPheLysArgLeuGlnGlu	707
QY	2208	GAGGAGCTAAAAGTCGTGATTTACAGCCCGAGGTGCAGGCTCTCAAGGAGAAGATCCAC	2267
Db	708	GluGluAlaLysSerGlyHisLeuSerArgGluValAspAlaLeuLysGluLysIleHis	727
QY	2268	GAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTATTCGGTCCCTTCAG	2327
Db	728	GluTyrMetAlaThrGluAspLeuIleCysHisLeuGlnGlyAspHisSerValLeuGln	747
QY	2328	CAAAGATTATGGAAGAAGAAACTAGAAACAAGAACATGGGAGGGAGGTCTCTCAATCTG	2387
Db	748	LysLysLeuAsnGlnGlnGluAsnArgAsnArgAspLeuGlyArgGluIleGluAsnLeu	767
QY	2388	ACCAAGGAGCTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGCCGAGTGGGAACGGC	2447
Db	768	ThrLysGluLeuGluArgTyrArgHisPheSerLysSerLeuArgProSerLeuAsnGly	787
QY	2448	CGAAGGATGGTGGACGTGCTGTGGCTCCACTGGGTGCAGACCGAGCGGTGTGCGGG	2507
Db	788	ArgArgIleSerAspProGlnValPheSerLysGluValGlnThrGluAlaVal-----	805
QY	2508	GATGCTCGGAGGAGGAGACCCCG-----GCTGTG	2537
Db	806	-----AspAsnGluProProAspTyrLysSerLeuIleProLeuGluArgAlaVal	822
QY	2538	TTCATTCGCAAAATCCTTCCAGGAG---GAAAATCACATCATGAGTAATCTTCGACAGGTA	2594
Db	823	IleAsnGlyGlnLeuTyrGluGluSerGluAsnGlnAsp-----	835
QY	2595	GGCCTGAAGAAACCCATGGAAACGGTCTCTCGTCCCTC-----GACAGGTAT	2639
Db	836	-----GluAspProAsnAspGluGlySerValLeuSerPheLysCysSerGlnSerThr	853
QY	2640	CCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTGATTCCTTGGATGAGAAAAAGA	2699
Db	854	ProCysProValAsn-----ArgLysLeuTrpIleProTrpMetLysSerLys	869
QY	2700	GAA-----AACGGTCCCTTCCACTCCGCGAGGAGAAAGGGCCCGAGGCCCCAACCCAG	2747
Db	870	GluGlyHisLeuGlnAsnGly-----LysMetGlnThrLysProAsnAla	884
QY	2748	GGTCAGGGCACCCCGGGAGCTGGTCTCTAGCACCAAAAGCAGGGCCAGCCCTACACATC	2807
Db	885	AsnPheValGlnProGlyAspLeuValLeuSerHisThrProGlyGlnProLeuHisIle	904
QY	2808	CGTGTGACACCAAGATCATGAGAACAGCACTGCCACCCCTGGAGATCACAGCCCCACATCT	2867
Db	905	LysValThrProAspHisValGlnAsnThrAlaThrLeuGluIleThrSerProThrThr	924
QY	2868	GAA-----GAGTTTCTCTAGTACCACCGCTCATCTTACCTTAGGCAACCCAGAAACCA	2921
Db	925	GluSerProHisSerTyrThrSerThrAlaValIleProAsnCysGlyThrProLysGln	944
QY	2922	AGATAACCATTTATCCATCACCCAATGTATGTGCGAAAAAGCCCCAAA---AGTCAGAT	2978
Db	945	ArgIleThrIleLeuGlnAsnAlaSerIleThrProValLysSerLysThrSerThrGlu	964

QY 2979 CCTACTCTCGGCCAGAGCCAGCATGTCCCTGTGCACGATTACTACTATTTCAGAGAG 3038
Db 965 AspLeuMetAsnLeuGluGlnGlyMetSerProIleThrMetAlaThrPheAlaArgAla 984
QY 3039 AAGAGCCCGGAAGGTGGAAGGAGCGCCTTTGCCGACAGGCCTGCATCCCCCATCCAAATC 3098
Db 985 GlnThrProGluSerCysGlySerLeuThrProGluArgThrMetSerProIleGlnVal 1004
QY 3099 ATGACGGTGTCAACATCTGCAGCTCCCACCTGAATCGCTGTCTCTCCTGAATCTCAGGAA 3158
Db 1005 LeuAlaValThrGlySerAlaSerSerProGluGlnGlyArgSerProGluProThrGlu 1024
QY 3159 GTGCCTATGGGAAGGACTATCCTCAAAGTCACCCCGGAAAAACAAACTGTTCCAGCCCCC 3218
Db 1025 IleSerAlaLysHisAlaIlePheArgValSerProAspArgGlnSerSerTrpGlnPhe 1044
QY 3219 GTGCGGAAGTACAACCTCAATGCTAATATCATCACGACGGAAGACAAATAAATTCACATT 3278
Db 1045 GlnArgSerAsnSerAsnSerSerValIleThrThrGluAspAsnLysIleHisIle 1064
QY 3279 CACCTGGGTTCTCAGTTTAAGCGATCTCCTGGSCCTCGCGCTGAAGCGGTGAGCCAGTT 3338
Db 1065 HisLeuGlySerProTyrMetGlnAla-----ValAlaSerProValArgProAla 1081
QY 3339 ATCACCGTCCGGCCTGTCAACGTGACAGCGGAGAGGAGGTTTCTACAGGCACAGTCCTT 3398
Db 1082 SerProSerAlaProLeuGlnAspAsnArgThrGlnGlyLeuIleAsnGlyAlaLeuAsn 1101
QY 3399 CGCTCTCCAGGAACCACTCTCTTCAAGACCGGCTGCTAGCAAAAGTGACCAAGCACTATA 3458
Db 1102 LysThr-----ThrAsnLysValThrSerSerIle 1111
QY 3459 ACTATAACCCCGTGCACAACTCATCCACACGAGGAACCCAA 3500
Db 1112 ThrIleThrProThrAlaThrProLeuProArgGlnSerGln 1125

RESULT 4
US-10-309-851-12
; Sequence 12, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
; CURRENT APPLICATION NUMBER: US/10/309,851
; CURRENT FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1135
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-851-12

Alignment Scores:
Pred. No.: 5.87e-139 Length: 1135
Score: 2444.50 Matches: 541
Percent Similarity: 64.31% Conservative: 214
Best Local Similarity: 46.08% Mismatches: 338
Query Match: 31.62% Indels: 81
DB: 14 Gaps: 19

US-10-788-793-1 (1-4364) x US-10-309-851-12 (1-1135)

QY 75 ATGAGATCACGAATCAAGGTGGAGAAAGTTTCATCTAACGGGCATGTCTCCTGCCCCCAAG 134
Db 1 MetArgSerArg-----GlySerAspThrGluGlySerAlaGlnLysLysPheProArg 18
QY 135 TCCTCCATCATCAGCAGTGATGGTGAAGGGGCCCTCAGAGATGCAAAAAGAACAAAG 194

Db 19 HisThr-----LysGlyHisSerPheGlnGlyProLysAsnMet 31
QY 195 GCCAATCGGAAGGAGAG---GATGTCTATGGCTTCCGGAACCTATCAAAAGGCACCTCAAA 251
Db 32 LysHisArgGlnGlnAspLysAspSerProSerGluSerAspVal-----IleLeu 48
QY 252 CCATCTGGAGAAAGTGAGAAA-----AAGACTAAGAAAGTCTGTGGAGTTATCTC 299
Db 49 ProCysProLysAlaGluLysProHisSerGlyAsnGlyHisGlnAlaGluAspLeuSer 68
QY 300 AAGGAGGACCTCATCCAGCTCCTGAGTATCATTGAAGGGGAGTTTGCAGGCTCGAGAAGAT 359
Db 69 ArgAspAspLeuLeuPheLeuLeuSerIleLeuGluGlyGluLeuGlnAlaArgAspGlu 88
QY 360 GTCATCCACATGCTGAGGACAGAGAAAAACCAAGCCCCGAGGTTCTTGGAGGCACACTATGGA 419
Db 89 ValIleGlyIleLeuLysAlaGluLysMetAspLeuAlaLeuLeuGluAlaGlnTyrGly 108
QY 420 TCTGCAGAACCTGAGAAAAGTGCTTCCGGGTCTCTGACCGAGATGCCATCCTTGTCTCAAGAG 479
Db 109 PheValThrProLysLysValLeuGluAlaLeuGlnArgAspAlaPheGlnAlaLysSer 128
QY 480 AAGTCCATAGGAGAAGACGTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAA 539
Db 129 ThrProTrpGlnGluAspIleTyrArgArgIleLeuGlyGlnLeuLeuValAlaGluLysSerArg 168
QY 540 AAGCAGAAAGGAGACGTACCGCGCATGCTAGAGCAGCTGTCTGGTGTGAGAAGTGTCTAC 599
Db 149 LysHisLysGluSerTyrArgArgIleLeuGlyGlnLeuLeuValAlaGluLysSerArg 168
QY 600 AGGCGCACCGTGTACGAGCTGGAGAACGAGAAGCACAAGCACACTGACTACATGAACAAG 659
Db 169 ArgGlnThrIleLeuGluLeuGluGluGluLysArgLysHisLysGluTyrMetGluLys 188
QY 660 AGCGACGACTTCACCAACCTGTCTGGAGCAGGAGCAGAGAGGTTGAAAAAGCTCCTTGAA 719
Db 189 SerAspGluPheIleCysLeuLeuGluGlnGluCysGluArgLeuLysLeuIleAsp 208
QY 720 CAAGAAAAAGCTTACCAAGCCCGCAAGAAAGAAACGCTAAGCGGCTCAACAACTT 779
Db 209 GlnGluIleLysSerGlnGluGluLysGluGlnGluLysGluLysArgValThrThrLeu 228
QY 780 CGAGATGAGCTTGTGAAGCTCAAGTCCCTTCGCCCTCATGTTGGTGACGAGGCGAGATG 839
Db 229 LysGluGluLeuThrLysLeuLysSerPheAlaLeuMetValValAspGluGlnGlnArg 248
QY 840 CACATCGAGCAACTGGCCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAAGCTGAGG 899
Db 249 LeuThrAlaGlnLeuThrLeuGlnArgGlnLysIleGlnGluLeuThrThrAsnAlaLys 268
QY 900 GAGGAGGAAGAAAAACTCAAAGCGGTCACTTACAAATCCAAGGAAGACCGCCAGAAAGCTG 959
Db 269 GluThrHisThrLysLeuAlaLeuAlaGluAlaArgValGlnGluGluGlnLysAla 288
QY 960 CTCAGTTAGAAGTGGACTTCGAACACAGGCCCTCGAGGTTTCCAGGAGCACGAAGAG 1019
Db 289 ThrArgLeuGluLysGluLeuGlnThrGlnThrThrLysPheHisGlnAspGlnAspThr 308
QY 1020 ATGAACGCCAAATTTGGCGAATCAAGAATCTCAACACCGGCAACTTCGACTCAAACTGGTT 1079
Db 309 IleMetAlaLysLeuThrAsnGluAspSerGlnAsnArgGlnLeuGlnGlnLysLeuAla 328
QY 1080 GGCTTATCGCAAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAGGCAGAG 1139
Db 329 AlaLeuSerArgGlnIleAspGluLeuGluGluThrAsnArgSerLeuArgLysAlaGlu 348
QY 1140 GAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAAACTCCAGTCTC 1199
Db 349 GluGluLeuGlnAspIleLysGluLysIleSerLysGlyGluTyrGlyAsnAlaGlyIle 368
QY 1200 ATGGCGGAAGTCGAGAGTCTGCGCAAGCCGCTGTGAGATGGAGGCGCAAGGATGAAGAG 1259
Db 369 MetAlaGluValGluGluLeuArgLysArgValLeuAspMetGluGlyLysAspGluGlu 388

Db 1082 SerProSerAlaProLeuGlnAspAsnArgThrGlnGlyLeuIleAsnGlyAlaLeuAsn 1101
QY 3399 CGTCTCCAGGAACCACTCTCTTCAAGACCCGGTGTAGCAAAAGTGACCACTATA 3458
Db 1102 LysThr-----ThrAsnLysValThrSerSerIle 1111
QY 3459 ACTATAACCCCGGTCAACAGTCATCCACACGAGGAACCCAA 3500
Db 1112 ThrIleThrProThrAlaThrProLeuProArgGlnSerGln 1125

RESULT 5
US-10-108-260A-4080
; Sequence 4080, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4080
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-4080

Alignment Scores:
Pred. No.: 9.48e-137 Length: 512
Score: 2406.50 Matches: 482
Percent Similarity: 97.66% Conservative: 18
Best Local Similarity: 94.14% Mismatches: 11
Query Match: 31.13% Indels: 1
DB: 15 Gaps: 1

US-10-788-793-1 (1-4364) x US-10-108-260A-4080 (1-512)

QY 75 ATGAGATCACGAAATCAAGGTGGAGAAAGTTTCATCTAAACGGGCATGTCTCTCTGCCCAAG 134
Db 1 MetArgSerArgAsnGlnGlyGlyGluSerAlaSerAspGlyHisIleSerCysProLys 20
QY 135 TCCTCCATCATCAGCAGTGATGGTGGTAAGGGCCCTCAGAAGATGCA---AAAAAGAAC 191
Db 21 ProSerIleIleGlyAsnAlaGlyGlyLysSerLeuSerGluAspAlaLysLysLys 40
QY 192 AAGGCCAATCGGAAGGAGGAGGATGTCTATGGCTTCCGGAACCTATCAAAGGCACTCAA 251
Db 41 LysSerAsnArgLysGluAspAspValMetAlaSerGlyThrValLysArgHisLeuLys 60
QY 252 CCATCTGGAGAAAGTGAGAAAAAGACTAAGAAAGTCTGTGGAGTTATCCAAGGAGACCTC 311
Db 61 ThrSerGlyGluCysGluArgLysThrLysLysSerLeuGluLeuSerLysGluAspLeu 80
QY 312 ATCCAGCTCCTGAGTATCATGGAAGGGGAGTTGCAGSCTCGAGAAGATGTCATCCACATG 371
Db 81 IleGlnLeuLeuSerIleMetGluGlyGluLeuGlnAlaArgGluAspValIleHisMet 100
QY 372 CTGAGGACAGAGAAAACCAAGCCCGAGGTTCTGGAGGCACACTATGGATCTGCAGAACCT 431
Db 101 LeuLysThrGluLysThrLysProGluValLeuGluAlaHisTyrGlySerAlaGluPro 120
QY 432 GAGAAAGTGTCTCGGTCTGCACCGAGATGCCATCCTTGCTCAAGAGAAGTCCATAGGA 491
Db 121 GluLysValLeuArgValIleuHisArgAspAlaIleLeuAlaGlnGluLysSerIleGly 140
QY 492 GAAGACGTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAAAAAGCAGAAAGGAG 551
Db 141 GluAspValTyrGluLysProIleSerGluLeuAspArgLeuGluGluLysGlnLysGlu 160
QY 552 ACGTACCGCCGCATGCTAGACGAGCTGCTGCTGGCTGAGAAGTGTACAGGGCGCACCGGTG 611

Db 161 ThrTyrArgArgMetLeuGluGlnLeuLeuLeuAlaGluLysCysHisArgArgThrVal 180
QY 612 TACGAGCTGGAGAACGAGAGACACAAGCACACTGACTACATGAACAAGAGCGACTTC 671
Db 181 TyrGluLeuGluAsnGluLysHisLysHisThrAspTyrMetAsnLysSerAspAspPhe 200
QY 672 ACCAACCTGCTGGAGCAGGAGCGAGAGAGGTGAAAAAGCTCCTTGAACAAGAAAAAGCT 731
Db 201 ThrAsnLeuLeuGluGlnGluArgGluArgLeuLysLysLeuLeuGluGlnGluLysAla 220
QY 732 TACCAAGCCCGCAAGAAAAAGGAAACCGCTAAGCGGCTCAACAACTTCGAGATGAGCTT 791
Db 221 TyrGlnAlaArgLysGluLysGluAsnAlaLysArgLeuAsnLysLeuArgAspGluLeu 240
QY 792 GTGAAGCTCAAGTCTCTTCGCCCTCATGTTGTGGACGAGAGGCAGATGCACATCGAGCAA 851
Db 241 VallysLeuLysSerPheAlaLeuMetLeuValAspGluArgGlnMetHisIleGluGln 260
QY 852 CTGGCCCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAGCTGAGGGAGGAGGAAGAA 911
Db 261 LeuGlyLeuGlnSerGlnLysValGlnAspLeuThrGlnLysLeuArgGluGluGluGlu 280
QY 912 AAACCTCAAAGCGGTCACTTACAAATCCAAGAAAGACCGCCAGAAAGTGTCTCAAGTTAGAA 971
Db 281 LysLeuLysAlaIleThrSerLysSerLysGluAspArgGlnLysLeuLeuLysLeuGlu 300
QY 972 GTGGACTTCGAAACAAGGCCCTCGAGGTTTCCAGGAGCACGAAGAGATGAACGCCAAA 1031
Db 301 ValAspPheGluHisLysAlaSerArgPheSerGlnGluHisGluGluMetAsnAlaLys 320
QY 1032 TTGGCGAATCAAGAATCTCACAAACCCGCCAACTTCGACTCAAACCTGGTTGGCTTATCGCAA 1091
Db 321 LeuAlaAsnGlnGluSerHisAsnArgGlnLeuArgLeuLysLeuValGlyLeuThrGln 340
QY 1092 AGGATTGAGGAGCTGGAAGAGACCAATAAAGCCTTCAGAAGGCAGAGGAAGAGCTCCAG 1151
Db 341 ArgIleGluGluLeuGluGluThrAsnLysAsnLeuGlnLysAlaGluGluLeuGln 360
QY 1152 GAGCTGAGAGAGAAATTCCAAAGGGGAATGTGGAAACTCCAGTCTCATGGCGAAGTG 1211
Db 361 GluLeuArgAspLysIleAlaLysGlyGluCysGlyAsnSerSerLeuMetAlaGluVal 380
QY 1212 GAGAGTCTGCGAAGCGCGTGTGATGAGATGGAGGGCAAGGATGAAGAGATCACGAAGACC 1271
Db 381 GluAsnLeuArgLysArgValLeuGluMetGluGlyLysAspGluGluIleThrLysThr 400
QY 1272 GAGGCCAGTCCCGGAGCTGAAGAAGAAGTCCAAAGAGGAACACACACGAAGGAA 1331
Db 401 GluSerGlnCysArgGluLeuArgLysLysLeuGlnGluGluGluHisSerLysGlu 420
QY 1332 CTTAGACTAGAACTGGAGAAAGCTGCAGAAAGAGGATGTCTGAGCTGGAGAAAGCTGGAGAA 1391
Db 421 PheArgLeuGluValGluLysLeuGlnLysArgMetSerGluLeuGluLysLeuGluGlu 440
QY 1392 GCGTTCAGCCGAGTAAGTCGGAATGCCACCCAGCTCCATCTGAACCTGGAGAAAGGAGAA 1451
Db 441 AlaPheSerLysSerLysSerGluCysThrGlnLeuHisLeuAsnLeuGluLysGluLys 460
QY 1452 AACCTAACCAAGACCTGTGAAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTC 1511
Db 461 AsnLeuThrLysAspLeuLeuAsnGluLeuGluValValLysSerArgValLysGluLeu 480
QY 1512 GAATGCTCCGAGAGTAGACTGGAGAAGGCCGAGTTAAGCCTCAAAGATGACCTTACAAAG 1571
Db 481 GluCysSerGluSerArgLeuGluLysAlaGluLeuSerLeuLysAspAspLeuThrLys 500
QY 1572 CTGAAGTCTTCACTGTGATGCTGGTGGATGAGAGG 1607
Db 501 LeuLysSerPheThrValMetLeuValAspGluArg 512

RESULT 6
US-10-309-851-10
; Sequence 10, Application US/10309851

```

; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
; CURRENT APPLICATION NUMBER: US/10/309,851
; CURRENT FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 764
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-851-10

Alignment Scores:
Pred. No.:      8.38e-102      Length:      764
Score:          1828.00      Matches:     386
Percent Similarity: 69.55%    Conservative: 144
Best Local Similarity: 50.66%  Mismatches:  210
Query Match:    23.65%      Indels:      22
DB:             14          Gaps:         7

US-10-788-793-1 (1-4364) x US-10-309-851-10 (1-764)

QY      75  ATGAGATCAGGAATCAAGGTGGAGAAAGTTCATCTAACGGGCATGTCCTCTGCCCAAG 134
      |||||||
Db      1  MetArgSerArg-----GlySerAspThrGluGlySerAlaGlnLysLysPheProArg 18

QY     135  TCCTCCATCATCAGCAGTGATGGTGGTAAGGGCCCCCTCAGAAGATGCAAAAAAGAACAAG 194
      ::|||
Db     19  HisThr-----LysGlyHisSerPheGlnGlyProLysAsnMet 31

QY     195  GCCAATCGGAAGGAGGAG--GATGTCAATGGCTTCGGGAACCTATCAAAAGGCACCTCAAA 251
      ::|||
Db     32  LysHisArgGlnGlnAspLysAspSerProSerGluSerAspVal-----IleLeu 48

QY     252  CCATCTGGAGAAAGTGAGAAA-----AAGACTAAGAAGTCTGTGGAGTTATCC 299
      |||
Db     49  ProCysProLysAlaGluLysProHisSerGlyAsnGlyHisGlnAlaGluAspLeuSer 68

QY     300  AAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGAGTTGCAGGCTCGAGAAAGT 359
      ::|||
Db     69  ArgAspAspLeuLeuPheLeuLeuSerIleLeuGluGlyGluLeuGlnAlaArgAspGlu 88

QY     360  GTCAATCCACATGCTGAGGCACAGAGAAAACCAAGCCCGAGGTTCTGGAGGCACACTATGGA 419
      |||||
Db     89  ValIleGlyIleLeuLysAlaGluLysMetAspLeuAlaLeuLeuGluAlaGlnTyrGly 108

QY     420  TCTGCAGAACCTGAGAAAGTGTCTTCGGGTCTGCACCGAGATGCCATCCTTGTCTCAAGAG 479
      |||||
Db    109  PheValThrProLysLysValLeuGluAlaLeuGlnArgAspAlaPheGlnAlaLysSer 128

QY     480  AAGTCCATAGGAGAAGACGTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAA 539
      |||||
Db    129  ThrProTrpGlnGluAspIleTyrGluLysProMetAsnGluLeuAspLysValValGlu 148

QY     540  AAGCAGAAAGGAGACGTACCGCCGCATGCTAGAGCAGCTGCTGGCTGAGAAAGTGTCTAC 599
      |||||
Db    149  LysHisLysGluSerTyrArgArgIleLeuGlyGlnLeuLeuValAlaGluLysSerArg 168

QY     600  AGGCGCACCGTGTACGAGTGGAGAACGAGAAGCACAAGCACACTACTACTGAACAAG 659
      |||||
Db    169  ArgGlnThrIleLeuGluLeuGluGluGluLysArgLysHisLysGluTyrMetGluLys 188

QY     660  AGCGACGACTTCACCAACCTGTGGAGCAGGAGCGAGAGAGGTTGAAAAAGCTCCCTTGA 719
      |||||
Db    189  SerAspGluPheIleCysLeuLeuGluGlnGluCysGluArgLeuLysLysLeuIleAsp 208

QY     720  CAAGAAAAAGCTTACCAGCCCGCAAGAAAAAGGAAACGCTAAGCGGCTCAACAACTT 779
      |||||

```

Db	209	GlnGluIleLysSerGlnGluGluLysGluGlnGluLysGluLysArgValThrThrLeu	228						
QY	780	CGAGATGAGCTTGTGAAGCTCAAGTCTCTCGCCCTCATGTTGGTGGACGAGAGGCAGATG	839						
Db	229	LysGluGluLeuThrLysLeuLysSerPheAlaLeuMetValValAspGluGlnArg	248						
QY	840	CACATCGAGCAACTGGGCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAGCTGAGG	899						
Db	249	LeuThrAlaGlnLeuThrLeuGlnArgGlnLysIleGlnGluLeuThrThrAsnAlaLys	268						
QY	900	GAGGAGGAAGAAAAAATCAAAAGCGGTCACTTACAAATCCAAGGAAGACCGCCAGAAAGCTG	959						
Db	269	GluThrHisThrLysLeuAlaLeuAlaGluAlaArgValGlnGluGluGlnLysAla	288						
QY	960	CTCAAGTTAGAAGTGGACTTCGAACACAAAGGCCTCGAGGTTTCCAGGAGCACGAAGAG	1019						
Db	289	ThrArgLeuGluLysGluLeuGlnThrGlnThrThrLysPheHisGlnAspGlnAspThr	308						
QY	1020	ATGAACGCCAAATTCGCGAATCAAGAAATCTCAACACCGCAACTTCGACTCAAACTGGTT	1079						
Db	309	IleMetAlaLysLeuThrAsnGluAspSerGlnAsnArgGlnLeuGlnGlnLysLeuAla	328						
QY	1080	GGCTTATCGCAAGGATTGAGGAGCTGGAAGACAGACCATAAAAGCCTTCAGAAGGCAGAG	1139						
Db	329	AlaLeuSerArgGlnIleAspGluLeuGluThrAsnArgSerLeuArgLysAlaGlu	348						
QY	1140	GAAGAGCTCCAGGAGCTGAGAGAGAAAATTCCAAAGGGGAATGTGGAAACTCCAGTCTC	1199						
Db	349	GluGluLeuGlnAspIleLysGluLysIleSerLysGlyGluTyrGlyAsnAlaGlyIle	368						
QY	1200	ATGGCGAAAGTGGAGAGTCTGCGCAAGCGCGTGTTCAGATGGAGGGCAAGGATGAAGAG	1259						
Db	369	MetAlaGluValGluGluLeuArgLysArgValLeuAspMetGluGlyLysaspGluGlu	388						
QY	1260	ATCACGAAGACCGAGGCCCGAGTGCCGGGAGCTGAAGAAGAAGCTCCAAGAGGAAGAACAC	1319						
Db	389	LeuIleLysMetGluGluGlnCysArgAspLeuAsnLysArgLeuGluArgGluThrLeu	408						
QY	1320	CACAGCAAGGAACTTAGACTAGAAAGTGGAGAAAGCTGCAGAAAGAGGATGTCCTGAGCTGGAG	1379						
Db	409	GlnSerLysAspPheLysLeuGluValGluLysLeuSerLysArgIleMetAlaLeuGlu	428						
QY	1380	AAGCTGGAGGAAGCGTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTG	1439						
Db	429	LysLeuGluAspAlaPheAsnLysSerLysGlnGluCysTyrSerLeuLysCysAsnLeu	448						
QY	1440	GAGAAGGAGAAGAACCTTAACCAAGACCTGCTGTAACGAGCTGGAGGTGGTCAAGAGTCGA	1499						
Db	449	GluLysGluArgMetThrThrLysGlnLeuSerGlnGluLeuGluSerLeuLysValArg	468						
QY	1500	GTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAGGCCGAGTTAAGCCTCAAAGAT	1559						
Db	469	IleLysGluLeuGluAlaIleGluSerArgLeuGluLysThrGluPheThrLeuLysGlu	488						
QY	1560	GACCTTACAAAGCTGAAGTCCTTCACTGTGATGCTGCTGGATGAGAGGAAAAATATGATG	1619						
Db	489	AspLeuThrLysLeuLysThrLeuThrValMetPheValAspGluArgLysThrMetSer	508						
QY	1620	GAGAAAATAAGCAAGAAGAGAGGAAAGTGGATGGGTTGAATAAAACTTTAAGGTGGAG	1679						
Db	509	GluLysLeuLysLysThrGluAspLysLeuGlnAlaSerSerGlnLeuGlnValGlu	528						
QY	1680	CAGGAAAAAGTCATGGATGTGACGGAAAAAGCTAATCGAGGAAAGCAAGAGCTTTTAAAA	1739						
Db	529	GlnAsnLysValThrThrValThrGluLysLeuIleGluGluThrLysArgAlaLeuLys	548						
QY	1740	CTCAAAATCTGAAATCGAGGAAAAAGGAGTACAGTCTGACAAAAGGAGAGGATGAGCTGATG	1799						
Db	549	SerLysThrAspValGluGluLysMetTyrSerValThrLysGluArgAspAspLeuLys	568						
QY	1800	GGTAAACTGAGGAGCGAAGAAGAGGTCTCTGTGTAAGTCTGAGCTGAGCTGTAGACTTACTA	1859						
Db	569	AsnLysLeuLysAlaGluGluGluLysGlyAsnAspLeuLeuSerArgValAsnMetLeu	588						

QY 1860 AAGAAGCGGCTTGATGGCATAGAGGAGGTAGAAAGGGAATAAAACCGAGGTAGG----- 1913
Db LysAsnArgLeuGlnSerLeuGluAlaIleGluLysAspPheLeuLysAsnLysLeuAsn 608
QY 1914 -----TCGTGCAAGGGGCTGTGAGTTCACCTGCCCGGAAGACAATAAGATCAGAGAACTA 1967
Db GlnAspSerGlyLysSerThrAlaLeuHisGlnGluAsnAsnLysIleLysGluLeu 628
QY 1968 ACCTTTGAAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTGGTGGAGGGGAC 2027
Db SerGlnGluValGluArgLeuLysLeuLysLeuLysAspMetLysAlaIleGluAspAsp 648
QY 2028 TTGATGAAGACCGAGGACGAATATGACCAGTTGGAGCAGAAGTTTCAGAACCCGAGCAGGAT 2087
Db LeuMetLysThrGluAspGluTyrGluThrLeuGluArgArgTyrAlaAsnGluArgAsp 668
QY 2088 AAGGCAAACTTCCTCTCCAGCAGCTCGAGGAAATCAAAACACCAAAATGGCCAAAGCACAAA 2147
Db LysAlaGlnPheLeuSerLysGluLeuGluHisValLysMetGluLeuAlaLysTyrLys 688
QY 2148 GCCATAGAGAAAGGGGAGCGGTGAGCCAGGAAGCCGAACTCGCACACAGTTTCGGCTG 2207
Db LeuAlaGluLysThrGlu---ThrSerHisGluGlnTrpLeuPheLysArgLeuGlnGlu 707
QY 2208 GAGGAGGCTAAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAGGAGAGATCCAC 2267
Db GluGluAlaLysSerGlyHisLeuSerArgGluValAspAlaLeuLysGluLysIleHis 727
QY 2268 GAGCTGATGAACAAGGAGACCAAGCTGTCTCAGCTCAAAGTCGACTATTTCGGTCTTCAG 2327
Db GluTyrMetAlaThrGluAspLeuIleCysHisLeuGlnGlyAspHisSerValleuGln 747
QY 2328 CAAAGA 2333
Db LysLys 749

RESULT 7

US-10-108-260A-3569
; Sequence 3569, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3569
; LENGTH: 824
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3569

Alignment Scores:

Pred. No.: 5.97e-101 Length: 824
Score: 1814.00 Matches: 401
Percent Similarity: 64.67% Conservative: 150
Best Local Similarity: 47.07% Mismatches: 237
Query Match: 23.46% Indels: 64
DB: 15 Gaps: 14

US-10-788-793-1 (1-4364) x US-10-108-260A-3569 (1-824)

QY 1026 GCCAAATTGGCGAATCAAGAATCTCAACACCGGCAACTTCGACTCAAACCTGGTGGCTTA 1085
Db 2 AlaLysLeuThrAsnGluAspSerGlnAsnArgGlnLeuGlnGlnLysLeuAlaLeu 21
QY 1086 TCGCAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCCTTCAGAAGCGCAGAGGAAGAG 1145
Db 22 SerArgGlnIleAspGluLeuGluGluThrAsnArgSerLeuArgLysAlaGluGluGlu 41

QY 1146 CTCCAGGAGCTGACAGAGAAAAATTGCCAAAGGGGAANTGTGGAAACTCCAGTCTCATGGCG 1205
Db LeuGlnAspIleLysGluLysIleSerLysGlyTyrGlyAsnAlaGlyIleMetAla 61
QY 1206 GAAGTGGAGAGTCTCGCAAGCGCGTGTGAGATGGAGGGCAAGGATGAAGAGATCACG 1265
Db GluValGluGluLeuArgLysArgValLeuAspMetGluGlyLysAspGluGluIle 81
QY 1266 AAGACCGAGGCCCACTGCGGGAGCTGAAGAAGAAGCTCCAAGAGGAAGAACACACACAGC 1325
Db LysMetGluGluGlnCysArgAspLeuAsnLysArgLeuGluArgGluThrLeuGlnSer 101
QY 1326 AAGAACTTAGACTAGAAAGTGGAGAAGCTGCAGAAGAGGATGTCTGAGCTGGAGAAGCTG 1385
Db LysAspPheLysLeuGluValGluLysLeuSerLysArgIleMetAlaLeuGluLysLeu 121
QY 1386 GAGAAAGCGTTACGCCGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAG 1445
Db GluAspAlaPheAsnLysSerLysGlnGluCysTyrSerLeuLysCysAsnLeuGluLys 141
QY 1446 GAGAAGAACCTAACCAAGACCTGTGTAACGAGCTGGAGGTGTCGAAGAGTCGAGTTAAA 1505
Db GluArgMetThrThrLysGlnLeuSerGlnGluLeuGluSerLeuLysValArgIleLys 161
QY 1506 GAACTCGAATGCTCCGAGAGTAGACTGGAGAAGCCGAGTTAAGCCTCAAAGATGACCTT 1565
Db GluLeuGluAlaIleGluSerArgLeuGluLysThrGluPheThrLeuLysGluAspLeu 181
QY 1566 ACAAAGCTGAAGTCCTTCACTGTGATGCTGTGGATGAGAGGAAAAATATGATGGAGAAA 1625
Db ThrLysLeuLysThrLeuThrValMetPheValAspGluArgLysThrMetSerGluLys 201
QY 1626 ATAAAGCAAGAGAGAGAAAGTGGATGGGTGGAATAAAACCTTTAAGGTGGAGCAGGGA 1685
Db LeuLysLysThrGluAspLysLeuGlnAlaAlaSerSerGlnLeuGlnValGluGlnAsn 221
QY 1686 AAAGTCATGGATGTACCGGAAAAAGCTAATCGAGGAAGCAAGAGCTTTTAAAACTCAAA 1745
Db LysValThrThrValThrGluLysLeuIleGluGluThrLysArgAlaLeuLysSerLys 241
QY 1746 TCTGAAATGGAGGAAAAAGGAGTACAGTCTGACAAAGGAGAGGATGAGCTGATGGGTAAA 1805
Db ThrAspValGluGluLysMetTyrSerValThrLysGluArgAspLeuLysAsnLys 261
QY 1806 CTGAGGAGCGAAGAAAGGTCCTGTGTAACGTGAGCTGCAGTGTAGACTTACTAAAGAAG 1865
Db LeuLysAlaGluGluGluLysGlyAsnAspLeuLeuSerArgValAsnMetLeuLysAsn 281
QY 1866 CGGCTTGATGGCATAGAGGAGGTAGAAAGGGGAAATAAACCGAGGTAGG----- 1913
Db ArgLeuGlnSerLeuGluAlaIleGluLysAspPheLeuLysAsnLysLeuAsnGlnAsp 301
QY 1914 TCGTGCAAGGGTCTGAGTTCACCTGCCCGGAAGACAAATAAGATCAGAGAACTAACGCTT 1973
Db SerGlyLysSerThrThrAlaLeuHisGlnGluAsnAsnLysIleLysGluLeuSerGln 321
QY 1974 GAAATCGAGAGACTGAAGAAAACGGCTCCAGCAGTTGGAGGTGGAGGGGACTTGATG 2033
Db GluValGluArgLeuLysLeuLysLeuLysAspMetLysAlaIleGluAspAspLeuMet 341
QY 2034 AAGACCGAGGACGAATATGACCAGTTGGAGCAGAAGTTTCAGAACCCGAGCAGGATAAGGCA 2093
Db LysThrGluAspGluTyrGluThrLeuGluArgArgTyrAlaAsnGluArgAspLysAla 361
QY 2094 AACTTCTCTCCAGCAGCTCGAGGAAATCAAAACACCAAAATGGCCAAGCACAAAGCCATA 2153
Db GlnPheLeuSerLysGluLeuGluHisValLysMetGluLeuAlaLysTyrLysLeuAla 381
QY 2154 GAGAAAGGGGAGCGCGTGAGCCAGGAAAGCCGAACTCGGACACAGGTTTCGGCTGGAGGAG 2213
Db GluLysThrGlu---ThrSerHisGluGlnTrpLeuPheLysArgLeuGlnGluGlu 400
QY 2214 GCTAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAGGAGGAAGATCCACGAGCTG 2273

Db 121 GluLeuGluAlaIleGluSerArgLeuGluLysThrGluPheThrLeuLysGluAspLeu 140
QY 1566 ACAAAAGCTGAAGTCCTTCACTGTGATGCTGGTGGATGAGAGGAAAAATATGATGAGAGAA 1625
Db 141 ThrLysLeuLysThrLeuThrValMetPheValAspGluArgLysThrMetSerGluLys 160
QY 1626 ATAAAGCAAGAAGAGAGGAAAGTGGATGGGTTGAATAAAAACTTTAAGGTGGAGCAGGA 1685
Db 161 LeuLysLysThrGluAspLysLeuGlnAlaAlaSerSerGlnLeuGlnValGluGlnAsn 180
QY 1686 AAAGTCATGGATGTCGCGAAAAAGCTAATCGAGGAAACCAAGAAGCTTTTAAAACTCAAA 1745
Db 181 LysValThrThrValThrGluLysLeuIleGluThrLysArgAlaLeuLysSerLys 200
QY 1746 TCTGAAATGGAGGAAAGGAGTACAGTCTGACAAAGGAGAGGGATGAGCTGATCGGTAAA 1805
Db 201 ThrAspValGluGluLysMetTyrSerValThrLysGluArgAspAspLeuLysAsnLys 220
QY 1806 CTGAGGAGCGAAGAAAGGTCTCTGTGAACCTGAGCTGCAGTGTAGACTTACTTAAAGAAG 1865
Db 221 LeuLysAlaGluGluLysGlyAsnAspLeuLysSerArgValAsnMetLeuLysAsn 240
QY 1866 CGGCTTGATGGCATAGAGGAGGTAGAAAGGGAATAAACCCGAGGTAGG----- 1913
Db 241 ArgLeuGlnSerLeuGluAlaIleGluLysAspPheLeuLysAsnLysLeuAsnGlnAsp 260
QY 1914 TCGTGAAGGGGTCTGAGTTCACCTGCCCGGAAGACAATAAGATCAGAGAACTAACGCTT 1973
Db 261 SerGlyLysSerThrThrAlaLeuHisGlnGluAsnLysIleLysGluLeuSerGln 280
QY 1974 GAAATCGAGAGACTGAAGAAACGGCTCCAGCAGTGTGGAGGTGGAGGGGAGTGTATG 2033
Db 281 GluValGluArgLeuLysLeuLysLeuLysAspMetLysAlaIleGluAspAspLeuMet 300
QY 2034 AAGACCGAGGACGAATATGACCAAGTGGAGCAGAGTTGAGAACCGAGCAGGATAAGGCA 2093
Db 301 LysThrGluAspGluTyrGluThrLeuGluArgArgTyrAlaAsnGluArgAspLysAla 320
QY 2094 AACTTCCTCTCCAGCAGCTCGAGGAATCAAACACCAAAATGGCCAAAGCACAAAGCCATA 2153
Db 321 GlnPheLeuSerLysGluLeuGluHisValLysMetGluLeuAlaLysTyrLysLeuAla 340
QY 2154 GAGAAAGGGAGCGCTGAGCCAGGAAGCCGAACTCCGACACAGGTTTCGGCTGGAGGAG 2213
Db 341 GluLysThrGlu---ThrSerHisGluGlnTrpLeuPheLysArgLeuGlnGluGlu 359
QY 2214 GCTAAAAGTCGTGATTACAGGCCGAGGTGAGGTGAGGCTCTCAAGGAGAGATCCACGAGCTG 2273
Db 360 AlaLysSerGlyHisLeuSerArgGluValAspAlaLeuLysGluLysIleHisGluTyr 379
QY 2274 ATGAACAAGGAAGACCAAGTGTCTCAGTCCAAAGTCGACTATTTCGGTCTCCAGCAAGA 2333
Db 380 MetAlaThrGluAspLeuIleCysHisLeuGlnGlyAspHisSerValLeuGlnLysLys 399
QY 2334 TTTATGGAAGAAGAACTAAGAAACAAGACATGGGGAGGAGGTCTCTCAATCTGACCAAG 2393
Db 400 LeuAsnGlnGlnGluAsnArgAsnArgAspLeuGlyArgGluIleGluAsnLeuThrLys 419
QY 2394 GAGCTAGAGCTTTCCACGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAAGG 2453
Db 420 GluLeuGluArgTyrArgHisPheSerLysSerLeuArgProSerLeuAsnGlyArgArg 439
QY 2454 ATGGTGGACGTGCTGTGGCTCCACTGGGGTGCAGACCGAGCGGTGTGCGGGGATGCT 2513
Db 440 IleSerAspProGlnValPheSerLysGluValGlnThrGluAlaVal----- 455
QY 2514 GCGGAGGAGGAGACCCCG-----GCTGTGTTTCATT 2543
Db 456 ---AspAsnGluProProAspTyrLysSerLeuIleProLeuGluArgAlaValIleAsn 474
QY 2544 CGCAAATCCTTCCAGGAG---GAAAATCACATCATGAGTAATCTTCGACAGGTAGGCCTG 2600
::: ::::: |||

Db 475 GlyGlnLeuTyrGluGluSerGluAsnGlnAsp----- 485
QY 2601 AAGAAACCCATGGAACGGTCTCTCGGTCCTC-----GACAGGTATCCCCCA 2645
Db 486 GluAspProAsnAspGluGlySerValLeuSerPheLysCysSerGlnSerThrProCys 505
QY 2646 GCAGCGAATGAGCTCACCATGAGGAAGTCTTGGATTCTTGGATGAGAAAAAGAGAA--- 2702
Db 506 ProValAsn-----ArgLysLeuTrpIleProTrpMetLysSerLysGluGly 521
QY 2703 -----AACGGTCTTCCACTCCGACGAGGAAAAAGGGCCCGAGGCCAAACCCAGGTGCA 2753
Db 522 HisLeuGlnAsnGly-----LysMetGlnThrLysProAsnAlaAsnPhe 536
QY 2754 GGGCACCCCGGGAGCTGGTCTTAGCACCAAAGCAGGGCCAGCCCTACACATCCGTGTG 2813
Db 537 ValGlnProGlyAspLeuValLeuSerHisThrProGlyGlnProLeuHisIleLysVal 556
QY 2814 ACACCAGATCATGAGAACAGCAGCACTGCCACCCCTGGAGATCACAAAGCCCCACATCTGAA--- 2870
Db 557 ThrProAspHisValGlnAsnThrAlaThrLeuGluIleThrSerProThrThrGluSer 576
QY 2871 ---GAGTTTTTCTTAGTACCACCGTCACTTACCTTACCTTAGGCAACCCAGAAACCAAGAATA 2927
Db 577 ProHisSerTyrThrSerThrAlaValIleProAsnCysGlyThrProLysGlnArgIle 596
QY 2928 ACCATTATTCCATCACCCCAATGTCTATGTCTGCGAAAAAGCCCAA---AGTGCAGATCCTACT 2984
Db 597 ThrIleLeuGlnAsnAlaSerIleThrProValLysSerLysThrSerThrGluAspLeu 616
QY 2985 CTCGGCCCCAGAACGAGCCATGTCCCTGTCTACGATTACTACTATTTCCAGAGAGAAGAGC 3044
Db 617 MetAsnLeuGluGlnGlyMetSerProIleThrMetAlaThrPheAlaArgAlaGlnThr 636
QY 3045 CCGGAAGGTGGAAGAGCGCCTTTGCCGACAGGCCTGCATCCCCCATCCAATCATGACG 3104
Db 637 ProGluSerCysGlySerLeuThrProGluArgThrMetSerProIleGlnValLeuAla 656
QY 3105 GTGTCAACATCTGCAGCTCCCACCTGAAATCGCTGTCTCTCTGAATCTCAGGAAGTGCCT 3164
Db 657 ValThrGlySerAlaSerSerProGluGlnGlyArgSerProGluProThrGluIleSer 676
QY 3165 ATGGGAAGGACTATCTCAAAGTCAACCCCGGAAAAACAAACTGTTCCAGCCCCCGTGC GG 3224
Db 677 AlaLysHisAlaIlePheArgValSerProAspArgGlnSerSerTrpGlnPheGlnArg 696
QY 3225 AAGTACAACTCCAATGCTAATATCATCACACGGAAGACAAATAAAATTCACATTACCTTG 3284
Db 697 SerAsnSerAsnSerSerSerValIleThrThrGluAspAsnLysIleHisIleHisLeu 716
QY 3285 GGTTCCTCAGTTTAAGCGATCTCCTGGGCTGCGGCTGAAGGCGTGAGCCCGATTATCACC 3344
Db 717 GlySerProTyrMetGlnAla-----ValAlaSerProValArgProAlaSerPro 733
QY 3345 GTCCGGCCTGTCAACGTGACAGCGGAGAGGAGGTTTCTACAGGCACAGTCTTCGCTCT 3404
Db 734 SerAlaProLeuGlnAspAsnArgThrGlnGlyLeuIleAsnGlyAlaLeuAsnLysThr 753
QY 3405 CCCAGGAACCACTCTCTTCAAGACCCCGGTGCTAGCAAAAGTGACCAAGTATAACTATA 3464
Db 754 -----ThrAsnLysValThrSerSerIleThrIle 763
QY 3465 ACCCGGTCACAACGTTCATCCACACGAGGAACCCAA 3500
Db 764 ThrProThrAlaThrProLeuProArgGlnSerGln 775

RESULT 9
US-10-171-311-48
; Sequence 48, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan

```
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 752
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-48

Alignment Scores:
Pred. No.:      8.56e-91      Length:      752
Score:          1645.50      Matches:     363
Percent Similarity: 64.16%      Mismatches:  216
Best Local Similarity: 46.30%      Indels:       65
Query Match:    21.28%      Gaps:         13
DB:              14

US-10-788-793-1 (1-4364) x US-10-171-311-48 (1-752)

QY      816 ATGTTGGTCAGCAGAGGCGAGATGCACATCGAGCAACTGGCCCTGCAGAGTCAGAAAGTC 875
Db      1 MetValValAspGluGlnGlnArgLeuThrAlaGlnLeuThrLeuGlnArgGlnLysIle 20

QY      876 CAGGACCTCACTCAGAAAGCTGAGGAGGAGGAAGAAAACTCAAAGCGGTCACTTACAAA 935
Db      21 GlnGluLeuThrThrAsnAlaLysGluThrHisThrLysLeuAlaLeuAlaGluAlaArg 40

QY      936 TCCAAGGAAGACCGCCAGAGAGCTGCTCAAGTTAGAAGTGGACTTCGAACACAAAGCCCTCG 995
Db      41 ValGlnGluGluGlnLysAlaThrArgLeuGluLysGluLeuGlnThrGlnThrThr 60

QY      996 AGGTTTTCAGGAGCAGCAAGAGAGATGAACGCCCAAATTGGCGGAATCAAGAATCTCACAA 1055
Db      61 LysPheHisGlnAspGlnAspThrIleMetAlaLysLeuThrAsnGluAspSerGlnAsn 80

QY      1056 CGGCAACTCGACTCAAACTGGTTGGCTTATCGCAAAGGATTGAGGAGCTGGAAGAGACC 1115
Db      81 ArgGlnLeuGlnGlnLysLeuAlaAlaLeuSerArgGlnIleAspGluLeuGluGluThr 100

QY      1116 AATAAAAGCCTTCAGAAGGCAGAGGAAGAGCTCCAGGAGCTGAGAGAGAAAAATTGCCCAA 1175
Db      101 AsnArgSerLeuArgLysAlaGluGluGluLeuGlnAspIleLysGluLysIleSerLys 120

QY      1176 GGGGAATGTGAAACTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTGCTT 1235
Db      121 GlyGluTyrGlyAsnAlaGlyIleMetAlaGluVal----- 132

QY      1236 GAGATGGAGGGCAAGGATGAAGAGATCAGGAAGACCGAGGCCCCAGTGC CGGAGCTGAAG 1295
Db      133 -----GluGluLeuIleLysMetGluGluGlnCysArgAspLeuAsn 146

QY      1296 AAGAAGTCCAAGAGGAAGAACACCAAGCAAGGAACCTTAGACTAGAACTGGAGAGCTG 1355
Db      147 LysArgLeuGluArgGluThrLeuGlnSerLysAspPheLysLeuGluValGluLysLeu 166

QY      1356 CAGAAGAGGATGTCTGAGCTGGAGAAAGCTGGAGGAAGCGTTTCAGCCCGAGTAAGTCGAA 1415
```

```
Db      167 SerLysArgIleMetAlaLeuGluLysLeuGluAspAlaPheAsnLysSerLysGlnGlu 186
QY      1416 TGCACCCAGCTCCATCTGAACCTGGAGAGGAGAAAGAACCTTAACCAAGACCTGCTGAAC 1475
Db      187 CysTyrSerLeuLysCysAsnLeuGluLysGluArgMetThrThrLysGlnLeuSerGln 206
QY      1476 GAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAG 1535
Db      207 GluLeuGluSerLeuLysValArgIleLysGluLeuGluAlaIleGluSerArgLeuGlu 226
QY      1536 AAGGCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTCAAGTCTTCACTGTGATGCTG 1595
Db      227 LysThrGluPheThrLeuLysGluAspLeuThrLysLeuLysThrLeuThrValMetPhe 246
QY      1596 GTGGATGAGAGGAAAAATATATGATGGAGAAAAATAAGCAAGAAGAGAGGAAAGTGGATGG 1655
Db      247 ValAspGluArgLysThrMetSerGluLysLeuLysLysThrGluAspLysLeuGlnAla 266
QY      1656 TTGAATAAAAAACTTTAAGGTGGAGCAGGGAAGAAAGTCATGGATGTGACGGAAAAAGCTAATC 1715
Db      267 AlaSerSerGlnLeuGlnValGluGlnAsnLysValThrValThrGluLysLeuIle 286
QY      1716 GAGGAAAGCAAGAGCTTTTAAAACTCAAATCTGAAATCGAGAGGAAAAAGGAGTACAGTCTG 1775
Db      287 GluGluThrLysArgAlaLeuLysSerLysThrAspValGluGluLysMetTyrSerVal 306
QY      1776 ACAAAGGAGAGGATGAGCTGATGGGTAAACTGAGGAGCGGAAGAAAGGTCCTGTGAA 1835
Db      307 ThrLysGluArgAspAspLeuLysAsnLysLeuLysAlaGluGluLysGlyAsnAsp 326
QY      1836 CTGAGCTGCAGTGTAGACTTACTATAAGAACGGCTTCATGGCATAGAGGAGGTAGAAAGG 1895
Db      327 LeuLeuSerArgValAsnMetLeuLysAsnArgLeuGlnSerLeuGluAlaIleGluLys 346
QY      1896 GAAATAAAACCGAGGTAGG-----TCGTGCAAGGGGTCTGAGTTCACCTGCCCG 1943
Db      347 AspPheLeuLysAsnLysLeuAsnGlnAspSerGlyLysSerThrThrAlaLeuHisGln 366
QY      1944 GAAGACAATAAGATCAGAGAACTAACGCTTGAAATCGAGAGACTGAAGAAACGGCTCCAG 2003
Db      367 GluAsnAsnLysIleLysGluLeuSerGlnGluValGluArgLeuLysLeuLysLeuLys 386
QY      2004 CAGTTGGAGGTGGTGGAGGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAG 2063
Db      387 AspMetLysAlaIleGluAspAspLeuMetLysThrGluAspGluTyrGluThrLeuGlu 406
QY      2064 CAGAAGTTCAGAACCGAGCAGAGATAAGGCAAACTTCCTCTCCCAGCAGCTCGAGGAAATC 2123
Db      407 ArgArgTyrAlaAsnGluArgAspLysAlaGlnPheLeuSerLysGluLeuGluHisVal 426
QY      2124 AAACACCAAATGGCCAAAGCACAAAGCCATAGAGAAAGGGGAGGCCGTGAGCCAGGAAGCC 2183
Db      427 LysMetGluLeuAlaLysTyrLysLeuAlaGluLysThrGlu---ThrSerHisGluGln 445
QY      2184 GAACTGCGACACAGGTTTCGGCTGGAGGAGGCTAAAAGTCGTGATTACAGGCCGAGGTG 2243
Db      446 TrpLeuPheLysArgLeuGlnGluGluAlaLysSerGlyHisLeuSerArgGluVal 465
QY      2244 CAGGCTCTCAAGGAAAGATCCACGAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTC 2303
Db      466 AspAlaLeuLysGluLysIleHisGluTyrMetAlaThrGluAspLeuIleCysHisLeu 485
QY      2304 CAAGTCGACTATTCCGTCCTTCAGCAAAAGATTTATGGAAGAAGAACTAAGAACAAAGAAC 2363
Db      486 GlnGlyAspHisSerValCysLysLysLysLeuAsnGlnGlnGluAsnArgAsnArgAsp 505
QY      2364 ATGGGAGGGAGGTCTCAATCTGACCAAGGAGCTAGAGCTTTCCAAAGCGCTACAGCCGA 2423
Db      506 LeuGlyArgGluIleGluAsnLeuThrLysGluLeuGluArgTyrArgHisPheSerLys 525
QY      2424 GCTCTCAGGCCGAGCTGGGAACGGCCGAAGGATGTGTGGACGTGCCTGTGGCCTCCACTGGG 2483
```

Db 526 SerLeuArgProSerLeuAsnGlyArgArgIleSerAspProGlnValPheSerLysGlu 545
QY 2484 GTGCAGACCGAGCGGTGTCGGGGGATGCTGCGGAGGAGGACCCCG----- 2531
Db 546 ValGlnThrGluAlaVal-----AspAsnGluProProAspTyrLysSer 560
QY 2532 -----GCTGTGTTTCATTTCGCAAAATCCTTCCAGGAG---GAAAATCAC 2570
Db 561 LeuIleProLeuGluArgAlaValIleAsnGlyGlnLeuTyrGluGluSerGluAsnGln 580
QY 2571 ATCATGAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCATGGAACGGTCCCTCGTCCCTC 2630
Db 581 Asp-----GluAspProAsnAspGluGlySerValLeu 591
QY 2631 -----GACAGGTATCCCCCAGCAGCGGAATGAGCTCACCATGAGGAAGTCT 2675
Db 592 SerPheLysCysSerGlnSerThrProCysProValAsn-----ArgLysLeu 607
QY 2676 TGGATTCTTGGATGAGAAAAAGAGAA-----AACGGTCTCTCCACTCCCGCAG 2723
Db 608 TrpIleProTrpMetLysSerLysGluGlyHisLeuGlnAsnGly----- 622
QY 2724 GAGAAAGGGCCCGAGGCCAAACACAGGGTGCAGGGCACCCCGGGGAGCTGGTCTTACACCA 2783
Db 623 LysMetGlnThrLysProAsnAlaAsnPheValGlnProGlyAspLeuValLeuSerHis 642
QY 2784 AAGCAGGGCCAGCCCTACACATCCGTGTGACACCAAGATCATGAGAACAGCAGCTGCCACC 2843
Db 643 ThrProGlyGlnProLeuHisIleLysValThrProAspHisValGlnAsnThrAlaThr 662
QY 2844 CTGGAGATCACAAAGCCCCACATCTGAA-----GAGTTTTTCTCTAGTACCACCGTCATT 2897
Db 663 LeuGluIleThrSerProThrThrGluSerProHisSerTyrThrSerThrAlaValIle 682
QY 2898 CCTACCTTAGGCAACCAAGAACCAAGAATAACCATATTATTCATCACCCCAATGTCATGTCG 2957
Db 683 ProAsnCysGlyThrProLysGlnArgIleThrIleLeuGlnAsnAlaSerIleThrPro 702
QY 2958 CAAAAGCCCAAA---AGTGCAGATCCTACTCTCGGCCAGAACGAGCCATGTCCCTGTC 3014
Db 703 ValLysSerLysThrSerThrGluAspLeuMetAsnLeuGluGlnGlyMetSerProIle 722
QY 3015 ACGATTACTACTATTTCAGAGAGAAAGAGCCCGGAAGGTGGAAGGAGCGCCTTTGCCGAC 3074
Db 723 ThrMetAlaThrPheAlaArgAlaGlnThrProGluSerCysGlySerLeuThrProGlu 742
QY 3075 AGGCCTGCATCC 3086
Db 743 ArgThrMetSer 746

RESULT 10
US-09-925-302-495
; Sequence 495, Application US/09925302
; Patent No. US2002004941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 495
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (8)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (416)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-495
Alignment Scores:
Pred. No.: 1.16e-38 Length: 439
Score: 780.50 Matches: 186
Percent Similarity: 58.30% Conservative: 74
Best Local Similarity: 41.70% Mismatches: 139
Query Match: 10.10% Indels: 47
DB: 9 Gaps: 11
US-10-788-793-1 (1-4364) x US-09-925-302-495 (1-439)
QY 2067 AAGTTCAGAACCCGACGAGGATAAGGCCAAACTTCCTCTCCCAGCAGCTCGAGGAATCAAA 2126
Db 5 ArgTyrAla**GluArgAspLysAlaGlnPheLeuSerLysGluLeuGluHisValLys 24
QY 2127 CACCAATGGCCCAAGCACAAAGCCATAGAGAAAGGGGAGGCCGTGAGCCAGGAAGCCGAA 2186
Db 25 MetGluLeuAlaLysTyrLysLeuAlaGluLysThrGlu---ThrSerHisGluGlnTrp 43
QY 2187 CTGCGACACAGGTTTCGGCTGGAGGAGGCTAAAAGTCGTGATTTACAGGCCGAGGTGCAG 2246
Db 44 LeuPheLysArgLeuGlnGluGluAlaLysSerGlyHisLeuSerArgGluValAsp 63
QY 2247 GCTCTCAAGGAGAGATCCACGAGCTGATGAACAAGGAAGACCAAGCTGTCTCAGCTCCAA 2306
Db 64 AlaLeuLysGluLysIleHisGluTyrMetAlaThrGluAspLeuIleCysHisLeuGln 83
QY 2307 GTCGACTATTTCGGTCTTCAGCAAAAGATTATTGGAAGAAGAACTAAGAACACGAACATG 2366
Db 84 GlyAspHisSerValLeuGlnLysLysLeuAsnGlnGlnGluAsnArgAsnArgAspLeu 103
QY 2367 GGGAGGAGGTCCTCAATCTGACCAAGGAGTAGAGCTTTCCAAAGCGCTACAGCCGAGCT 2426
Db 104 GlyArgGluIleGluAsnLeuThrLysGluLeuGluArgTyrArgHisPheSerLysSer 123
QY 2427 CTCAGGCCGAGTGGGAACGGCCGAAGGATGGTGGACGTGCTGTGGCCTCCAGTGGGGTG 2486
Db 124 LeuArgProSerLeuAsnGlyArgArgIleSerAspProGlnValPheSerLysGluVal 143
QY 2487 CAGACCGAGCGGTGTCGGGGATGCTGCGGAGGAGGAGACCCCG----- 2531
Db 144 GlnThrGluAlaVal-----AspAsnGluProProAspTyrLysSerLeu 158
QY 2532 -----GCTGTGTTTCATTTCGCAAAATCCTTCCAGGAG---GAAAATCACATC 2573
Db 159 IleProLeuGluArgAlaValIleAsnGlyGlnLeuTyrGluGluSerGluAsnGlnAsp 178
QY 2574 ATGAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCATGGAACGGTCCCTCGTCCCTC--- 2630
Db 179 -----GluAspProAsnAspGluGlySerValLeuSer 189
QY 2631 -----GACAGGTATCCCCCAGCAGCGGAATGAGCTCACCATGAGGAAGTCTTGG 2678
Db 190 PheLysCysSerGlnSerThrProCysProValAsn-----ArgLysLeuTrp 205
QY 2679 ATTCCTTGGATGAGAAAAAGAGAA-----AACGGTCTTTCCTTCCACTCCGAGGAG 2726
Db 206 IleProTrpMetLysSerLysGluGlyHisLeuGlnAsnGly-----Lys 220
QY 2727 AAAGGGCCCGAGCCCAACACAGGGTGCAGGGCACCCTCCGGGAGCTGGTCTCTAGCACCAAAG 2786
Db 221 MetGlnThrLysProAsnAlaAsnPheValGlnProGlyAspLeuValLeuSerHisThr 240
QY 2787 CAGGGCCAGCCCTACACATCCGTGTGACACCAGATCATGAGAACAGCAGCTGCCACCCTG 2846
Db 241 ProGlyGlnProLeuHisIleLysValThrProAspHisValGlnAsnThrAlaThrLeu 260
QY 2847 GAGATCACAAAGCCCCACATCTGAA-----GAGTTTTTCTCTAGTACCACCGTCATTCCT 2900

Db 261 GluIleThrSerProThrThrGluSerProHisSerTyrThrSerThrAlaValIlePro 280
QY 2901 ACCTTAGGCAACACAGAAACCAAGAATAACCATATTATCCATCACCCCAATGTGCATGTGCGAA 2960
Db 281 AsnCysGlyThrProLysGlnArgIleThrIleLeuGlnAsnAlaSerIleThrProVal 300
QY 2961 AAGCCCAAA--AGTGCAGATCCTACTCTCGGCCCCAGAACAGCCATGTCCTCCCTGTCAAG 3017
Db 301 LysSerLysThrSerThrGluAspLeuMetAsnLeuGluGlnGlyMetSerProIleThr 320
QY 3018 ATTACTACTATTCCAGAGAGAAGAGCCCGGAAGGTGGAAGGAGCGCCTTTGCCGACAGG 3077
Db 321 MetAlaThrPheAlaArgAlaGlnThrProGluSerCysGlySerLeuThrProGluArg 340
QY 3078 CCTGCATCCCCCATCCAAATCATGACGGTGTCAACATCTGCAGCTCCCACTGAAATCGCT 3137
Db 341 ThrMetSerProIleGlnValLeuAlaValThrGlySerAlaSerSerProGluGlnGly 360
QY 3138 GTCTCTCCTGAATCTCAGGAAGTGCCTATGGGAAGGACTATCCTCAAAAGTCACCCCGGAA 3197
Db 361 ArgSerProGluProThrGluIleSerAlaLysHisAlaIlePheArgValSerProAsp 380
QY 3198 AAACAAACTGTTCCAGCCCCCGTCCGGAAGTACAACTCCCAATGCTAATATCATCACCCACG 3257
Db 381 ArgGlnSerSerTrpGlnPheGlnArgSerAsnSerAsnSerSerValIleThrThr 400
QY 3258 GAAGACAAATFAAAATTACATTCACCTGGGTTCTCAGTTTAAAGCGATCTCCTGGCCTGCC 3317
Db 401 GluAspAsnLysIleHisIleHisLeuGlySerProTyrMetGln**ValAlaSerPro 420
QY 3318 GCTGAAGGCGTGAGCCCA 3335
Db 421 ValArgProAlaSerPro 426

RESULT 11
US-09-925-302-495
; Sequence 495, Application US/09925302
; Publication No. US20030064072A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 495
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (416)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-495

Alignment Scores:
Pred. No.: 1.16e-38 Length: 439
Score: 780.50 Matches: 186
Percent Similarity: 58.30% Conservative: 74
Best Local Similarity: 41.70% Mismatches: 139
Query Match: 10.10% Indels: 47
DB: 12 Gaps: 11
US-10-788-793-1 (1-4364) x US-09-925-302-495 (1-439)

QY 2067 AAGTTCAGAACCGCAGCAGGATAAAGGCAAACTTCCTCTCCAGCAGCTCGAGGAATCAA 2126
Db 5 ArgTyrAla**GluArgAspLysAlaGlnPheLeuSerLysGluLeuGluHisValLys 24
QY 2127 CACCAAAATGGCCAAAGCACAAGCCATAGAGAAAGGGGAGGCCGCTGAGCCAGGAAGCCGAA 2186
Db 25 MetGluLeuAlaLysTyrLysLeuAlaGluLysThrGlu---ThrSerHisGluGlnTrp 43
QY 2187 CTGCGACACAGGTTTCGGCTGGAGGAGCTAAAGTCTGTGATTTACAGGCCGAGGTGCAG 2246
Db 44 LeuPheLysArgLeuGlnGluGluAlaLysSerGlyHisLeuSerArgGluValAsp 63
QY 2247 GCTCTCAAGGAGAAAGATCCACGAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTCCAA 2306
Db 64 AlaLeuLysGluLysIleHisGluTyrMetAlaThrGluAspLeuIleCysHisLeuGln 83
QY 2307 GTCGACTATTTCGGTCTTCAGCAAAAGATTTATGGAAGAAGAAACTAAGAAACAAGAACATG 2366
Db 84 GlyAspHisSerValLeuGlnLysLysLeuAsnGlnGlnGluAsnArgAsnArgAspLeu 103
QY 2367 GGGAGGAGGTCCTCAATCTGACCAAGAGCTAGAGCTTTCCAAGCGCTACAGCCGAGCT 2426
Db 104 GlyArgGluIleGluAsnLeuThrLysGluLeuGluArgTyrArgHisPheSerLysSer 123
QY 2427 CTCAGGCCGAGTGGGAACGCCGGAAGGATGGTGGACGTCGCTGTGGCTCTCCACTGGGGTG 2486
Db 124 LeuArgProSerLeuAsnGlyArgArgIleSerAspProGlnValPheSerLysGluVal 143
QY 2487 CAGACCGAGCGGTGTGCGGGGATGCTGCGGAGGAGGAGACCCCG----- 2531
Db 144 GlnThrGluAlaVal-----AspAsnGluProProAspTyrLysSerLeu 158
QY 2532 -----GCTGTGTTTCATTGCGAAATCCTTCCAGGAG---GAAAATCACATC 2573
Db 159 IleProLeuGluArgAlaValIleAsnGlyGlnLeuTyrGluGluSerGluAsnGlnAsp 178
QY 2574 ATGAGTAATCTTCGACAGGTAGGCTGGAAGAAACCCCATGGAACGGTCTCGGTCTCTC--- 2630
Db 179 -----GluAspProAsnAspGluGlySerValLeuSer 189
QY 2631 -----GACAGGTATCCCCAGCAGCAGCAATGAGCTACCATGAGGAAGTCTTG 2678
Db 190 PheLysCysSerGlnSerThrProCysProValAsn-----ArgLysLeuTrp 205
QY 2679 ATTCTCTGGATGAGAAAAAGAGAA-----AACGGTCTTCCACTCCGACGAGG 2726
Db 206 IleProTrpMetLysSerLysGluGlyHisLeuGlnAsnGly-----Lys 220
QY 2727 AAAGGGCCCGAGCCCAACACAGGGTGCAGGGCACCCCGGGGAGCTGGTCTTAGCACCAAAAG 2786
Db 221 MetGlnThrLysProAsnAlaAsnPheValGlnProGlyAspLeuValLeuSerHisThr 240
QY 2787 CAGGCCAGCCCTACACATCCGTGTGACACCAGATCATGAGAACAGACTGCCACCTG 2846
Db 241 ProGlyGlnProLeuHisIleLysValThrProAspHisValGlnAsnThrAlaThrLeu 260
QY 2847 GAGATCACAAAGCCCCACATCTGAA-----GAGTTTTTCTCTAGTACCACCGTCAATCCT 2900
Db 261 GluIleThrSerProThrThrGluSerProHisSerTyrThrSerThrAlaValIlePro 280
QY 2901 ACCTTAGGCAACCCAGAAACCAAGATAAACCATATTATCCATCACCCCAATGTATGCGCAA 2960
Db 281 AsnCysGlyThrProLysGlnArgIleThrIleLeuGlnAsnAlaSerIleThrProVal 300
QY 2961 AAGCCCAAA---AGTGCAGATCCTACTCTCGGCCCCAGAACGAGCCATGTCCCCTGTCAAG 3017
Db 301 LysSerLysThrSerThrGluAspLeuMetAsnLeuGluGlnGlyMetSerProIleThr 320
QY 3018 ATTACTACTATTTCAGAGAGAAGAGCCCGGAAGGTGGAAGGAGCGCCTTTGCCGACAGG 3077
Db 321 MetAlaThrPheAlaArgAlaGlnThrProGluSerCysGlySerLeuThrProGluArg 340

QY 3078 CCTGCATCCCCCATCCAAATCATGACGGTGTCAACATCTGCAGCTCCCACTGAAATCGCT 3137
Db 341 ThrMetSerProIleGlnValLeuAlaValThrGlySerAlaSerSerProGluGlnGly 360
QY 3138 GTCCTCTCTGAATCTCAGGAAGTGCCTATGGGAAGCACTATCCTCAAAGTCACCCCGAA 3197
Db 361 ArgSerProGluProThrGluIleSerAlaLysHisAlaIlePheArgValSerProAsp 380
QY 3198 AAACAAACTGTTCCAGCCCCCGTGGCGAAGTACAACCTCAATGCTAATATCATCACCAAG 3257
Db 381 ArgGlnSerSerTrpGlnPheGlnArgSerAsnSerAsnSerSerValIleThrThr 400
QY 3258 GAAGACAATAAAATTCACATTACCTGGGTTCTCAGTTTAAGCGATCTCCTGGGCTGCC 3317
Db 401 GluAspAsnLysIleHisIleHisLeuGlySerProTyrMetGln***ValAlaSerPro 420
QY 3318 GCTGAAGCGGTGAGCCCA 3335
Db 421 ValArgProAlaSerPro 426
RESULT 12
US-10-309-851-8
; Sequence 8, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
; CURRENT APPLICATION NUMBER: US/10/309,851
; CURRENT FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-851-8
Alignment Scores:
Pred. No.: 5,06e-38 Length: 350
Score: 769.50 Matches: 168
Percent Similarity: 66.02% Conservative: 71
Best Local Similarity: 46.41% Mismatches: 106
Query Match: 9.95% Indels: 17
DB: 14 Gaps: 5
US-10-788-793-1 (1-4364) x US-10-309-851-8 (1-350)
QY 75 ATGAGATCAGCAATCAAGGTGGAGAAAGTTTCATCTAACGGGCATGTCTCTGCCCCCAAG 134
Db 1 MetArgSerArg-----GlySerAspThrGluGlySerAlaGlnLysLysPheProArg 18
QY 135 TCCTCCATCATCAGCAGTGATGGTGAAGGGCCCTCAGAAGATGCAGAAAGAAAGAAACAAG 194
Db 19 HisThr-----LysGlyHisSerPheGlnGlyProLysAsnMet 31
QY 195 GCCAATCGGAAGGAGAG---GATGTCTATGGTTCGGAACTATCAAAAGGCACCTCAAA 251
Db 32 LysHisArgGlnGlnAspLysAspSerProSerGluSerAspVal-----IleLeu 48
QY 252 CCATCTGGAGAAAGTCAGAAA-----AAGACTAAGAAGTCTGTGGAGTTATCC 299
Db 49 ProCysProLysAlaGluLysProHisSerGlyAsnGlyHisGlnAlaGluAspLeuSer 68
QY 300 AAGGAGGACCTCATCCAGCTCCTCAGTATCATGGAAGGGGAGTTGCAGGTCGAGAAGAT 359
Db 69 ArgAspAspLeuPheLeuLeuSerIleLeuGluGlyGluLeuGlnAlaArgAspGlu 88
QY 360 GTCATCCACATGCTGAGGACAGAGAAACCAAGCCCGAGGTTCTGGAGGCACACTATGGA 419
Db 89 ValIleGlyIleLeuLysAlaGluLysMetAspLeuAlaLeuLeuGluAlaGlnTyrGly 108

QY 420 TCTGCAGAACCTGAGAAAAGTGTCTCGGTCCTGCACCGAGATGCCATCCTTGCTCAAGAG 479
Db 109 PheValThrProLysLysValLeuGluAlaLeuGlnArgAspAlaPheGlnAlaLysSer 128
QY 480 AAGTCCATAGGAGAAAGACGTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAA 539
Db 129 ThrProTrpGlnGluAspIleTyrGluLysProMetAsnGluLeuAspLysValValGlu 148
QY 540 AAGCAGAAGGAGACGTACCGCCGCATGCTAGAGCAGCTGCTGCTGGCTGAGAAGTGTAC 599
Db 149 LysHisLysGluSerTyrArgArgIleLeuGlyGlnLeuValAlaGluLysSerArg 168
QY 600 AGGCGCACCGTGTACGAGCTGGAGAACGAGAAGCACAAAGCACACTGACTACATGAACAAG 659
Db 169 ArgGlnThrIleLeuGluLeuGluGluGluLysArgLysHisLysGluTyrMetGluLys 188
QY 660 AGCGACGACTTCACCAACCTGCTGGAGCAGGAGCGAGAGAGGTTGAAAAGCTCCTTGAA 719
Db 189 SerAspGluPheIleCysLeuLeuGluGlnGluCysGluArgLeuLysLysLeuIleAsp 208
QY 720 CAAGAAAAGCTTACCAAGCCCGCAAGAAAAGAAAACGCTAAGCGGCTCAACAAACTT 779
Db 209 GlnGluIleLysSerGlnGluLysGluGlnGluLysGluLysArgValThrThrLeu 228
QY 780 CGAGATGAGCTTGTGAAGCTCAAGTCTTCGCCCCCTCATGTTGGTGGACGAGAGGCAGATG 839
Db 229 LysGluGluLeuThrLysLysLeuLysSerPheAlaLeuMetValValAspGluGlnGlnArg 248
QY 840 CACATCGAGCAACTGGGCTGCAGAGTCCAGAAAGTCCAGGACCTCACTCAGAAAGCTGAGG 899
Db 249 LeuThrAlaGlnLeuThrLeuGlnArgGlnLysIleGlnGluLeuThrThrAsnAlaLys 268
QY 900 GAGGAGGAAGAAAACCTCAAAGCGGTCACTTACAAATCCAAGGAAGACCGCCAGAAAGCTG 959
Db 269 GluThrHisThrLysLeuAlaLeuAlaGluAlaArgValGlnGlnGluGlnLysAla 288
QY 960 CTCAGATTAGAAGTGGACTTCGAACACAAAGGCCTCGAGGTTTCCAGGAGCAGCAAGAG 1019
Db 289 ThrArgLeuGluLysGluLeuGlnThrGlnThrThrLysPheHisGlnAspGlnAspThr 308
QY 1020 ATGAACGCCCAATTTGGCGAATCAAGAATCTCACAACCCGGCAACTTCGACTCAAACTGGT 1079
Db 309 IleMetAlaLysLeuThrAsnGluAspSerGlnAsnArgGlnLeuGlnGlnLysLeuAla 328
QY 1080 GGCTTATCGCAAGGATTGAGGAGTTCGAAAGAGACCAATAAAAGCCTTCAGAAAGGCAGAG 1139
Db 329 AlaLeuSerArgGlnIleAspGluLeuGluGluThrAsnArgSerLeuArgLysAlaGlu 348
QY 1140 GAAGAG 1145
Db 349 GluGlu 350
RESULT 13
US-10-309-851-6
; Sequence 6, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
; CURRENT APPLICATION NUMBER: US/10/309,851
; CURRENT FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-851-6

```
Alignment Scores:      1.8e-32      Length:      265
Pred. No.:            677.00      Matches:      138
Score:                70.94%      Conservative:  50
Percent Similarity:   52.08%      Mismatches:   77
Best Local Similarity: 8.76%      Indels:       0
Query Match:         14          Gaps:        0
DB:

US-10-788-793-1 (1-4364) x US-10-309-851-6 (1-265)

QY 351 CGAGAAAGATGTCATCCACATGCTGAGGACAGAGAAAACCAAGCCGAGGTTCTGGAGGCA 410
    |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 1 ArgAspGluValIleGlyIleLeuLysAlaGluLysMetAspLeuAlaLeuLeuGluAla 20

QY 411 CACTATGATCTGCAGAACCTTGAGAAAGTGCTTCGGGTCCTGCACCGAGATGCCATCCTT 470
    ||| ||| |||:||||| |||:||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21 GlnTyrGlyPheValThrProLysLysValLeuGluAlaLeuGlnArgAspAlaPheGln 40

QY 471 GCTCAAGAGAAAGTCCATAGGAGAAAGACGCTCTATGAGAAACCTATCTCAGAGCTGGACAGA 530
    |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 41 AlalysSerThrProTrpGlnGluAspIleTyrGluLysProMetAsnGluLeuAspLys 60

QY 531 CTGGAGGAAAAGCAGAAAGAGACGTACCGCCGATGCTAGAGCAGCTGCTGGCTGAG 590
    :||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 61 ValValGluLysHisLysGluSerTyrArgArgIleLeuGlyGlnLeuLeuValAlaGlu 80

QY 591 AAGTGTACAGGCGCACCGTGTACGAGCTGGAGAACGAGAACGACACACTGACTAC 650
    |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 81 LysSerArgArgGlnThrIleLeuGluLeuGluGluLysArgLysHisLysGluTyr 100

QY 651 ATGAACAAGAGCGACGACTTCACCAACCTGTCTGGAGCAGGAGCGAGAGGTTGAAAAG 710
    ||| ||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 101 MetGluLysSerAspGluPheIleCysLeuLeuGluGlnGluCysGluArgLeuLysLys 120

QY 711 CTCCTTGAACAAGAAAAGCTTACCAGCCCGCAAGAAAAGGAAAACGCTAAGCGGCTC 770
    |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 121 LeuIleAspGlnGluIleLysSerGlnGluGluLysGluGlnGluLysArgVal 140

QY 771 AACAACTTCGAGATGAGCTTGTGAAGCTCAAGTCTTCGCCCTCATGTTGGTGACGAG 830
    |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 141 ThrThrLeuLysGluGluLeuThrLysLeuLysSerPheAlaLeuMetValValAspGlu 160

QY 831 AGGCAGATGCACATCGAGCAACTGGGCGCTGCAGAGTCAAGAAAGTCCAGSACCTCACTCAG 890
    :||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 161 GlnGlnArgLeuThrAlaGlnLeuThrLeuGlnArgGlnLysIleGlnGluLeuThrThr 180

QY 891 AAGCTGAGGGAGGAGGAAGAAAACCTCAAAGCGGTCACTTACAAATCCAAGGAAGACCGC 950
    :||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 181 AsnAlalysGluThrHisThrLysLeuAlaLeuAlaGluAlaArgValGlnGluGlu 200

QY 951 CAGAAGCTGCTCAAGTTAGAAGTGGAATTGGAACACACAGGCGCTCGAGGTTTCCCAGGAG 1010
    ||| ||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 201 GlnLysAlaThrArgLeuGluLysGluLeuGlnThrGlnThrThrLysPheHisGlnAsp 220

QY 1011 CACGAAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCACAAACCGGCAACTTCGACTC 1070
    :||: ||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 221 GlnAspThrIleMetAlaLysLeuThrAsnGluAspSerGlnAsnArgGlnLeuGlnGln 240

QY 1071 AAACCTGGTGGCTTATCCGAAAAGATTGAGGAGCTTGGAGAGACCAATAAAAGCCTTCAG 1130
    ||| ||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 241 LysLeuAlaAlaLeuSerArgGlnIleAspGluLeuGluGluThrAsnArgSerLeuArg 260

QY 1131 AAGGCAGGAAGAG 1145
    ||| ||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 261 LysAlaGluGluGlu 265

RESULT 14
US-10-309-851-38
; Sequence 38, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
```

```
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
; CURRENT APPLICATION NUMBER: US/10/309,851
; CURRENT FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-851-38

Alignment Scores:      6.68e-29      Length:      384
Pred. No.:            618.50      Matches:      155
Score:                54.88%      Conservative:  70
Best Local Similarity: 37.80%      Mismatches:   126
Query Match:         8.00%      Indels:       59
DB:                14          Gaps:        12

US-10-788-793-1 (1-4364) x US-10-309-851-38 (1-384)

QY 2340 GAAGRAAGAACTAAGAACAAAGACATGGGGAGGGAGGTCTCTCAATCTGACCAAGGAGCTA 2399
    :||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 3 GlnGlnGluAsnArgAsnArgAspLeuGlyArgGluIleGluAsnLeuThrLysGluLeu 22

QY 2400 GAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAAGGATGGTG 2459
    ||| ||| |||:||||| |||:||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 23 GluArgTyrArgHisPheSerLysSerLeuArgProSerLeuAsnGlyArgArgIleSer 42

QY 2460 GACGTGCCTGTGGCTCCACTGGGGTGCAGACCGAGCGGTGTGCGGGGATGCTGCGGAG 2519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43 AspProGlnValPheSerLysGluValGlnThrGluAlaVal-----Asp 57

QY 2520 GAGGAGACCCCG-----GCTGTGTTCATTGCGCAA 2549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 58 AsnGluProProAspTyrLysSerLeuIleProLeuGluArgAlaValIleAsnGlyGln 77

QY 2550 TCCTTCCAGGAG---GAAAAATCACATCATGAGTAATCTTCGACAGGTAGGCCTGAAGAAA 2606
    ||| :||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 LeuTyrGluGluSerGluAsnGlnAsp-----GluAsp 88

QY 2607 CCCATGGAACGGTCTCGGTCTCTC-----GACAGGTATCCCCCAGCAGCG 2651
    ||| :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89 ProAsnAspGluGlySerValLeuSerPheLysCysSerGlnSerThrProCysProVal 108

QY 2652 AATGAGCTCACCATGAGGAAGTCTTGGATTCTCTGGATGAGAAAAAGAGAA----- 2702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109 Asn-----ArgLysLeuTrpIleProTrpMetLysSerLysGluGlyHisLeu 124

QY 2703 ---AACGGTCTTCCACTCCGACGAGAGAAAGGCCAGCCCAAACACGGGTGACGGGCAC 2759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 GlnAsnGly-----LysMetGlnThrLysProAsnAlaAsnPheValGln 139

QY 2760 CCCGGGAGCTGGTCTCTAGCACCAACAGCAGGGCCAGCCCTACACATCCGTGTGACACCA 2819
    ||| ||| :||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 140 ProGlyAspLeuValLeuSerHisThrProGlyGlnProLeuHisIleLysValThrPro 159

QY 2820 GATCATGAGAAACAGCACTGCCACCCCTGGAGATCACAGCCCCCACATCTGAA-----GAG 2873
    ||| ||| :||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 160 AspHisValGlnAsnThrAlaThrLeuGluIleThrSerProThrThrGluSerProHis 179

QY 2874 TTTTCTCTAGTACACCGTCACTTCCTACCTTAGGCAACCAAGAAACCAAGATAACCAATT 2933
    :||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 SerTyrThrSerThrAlaValIleProAsnCysGlyThrProLysGlnArgIleThrIle 199

QY 2934 ATTCCATCACCCCAATGTCATGTCGCAAAAGCCCAA---AGTGCAGATCCTACTCTCGGC 2990
    :||: ||| :||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 200 LeuGlnAsnAlaSerIleThrProValLysSerLysThrSerThrGluAspLeuMetAsn 219

QY 2991 CCAGAACGAGCCCATGTCCCTGTCCACGATTACTACTATTCCAGAGAGAGAGCCCGGAA 3050
    |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 220 LeuGluGlnGlyMetSerProIleThrMetAlaThrPheAlaArgAlaGlnThrProGlu 239
QY 3051 GGTGAAGGAGCGCCTTTCGGCAGACAGGCCTGCATCCCCCATCCAAATCATACGGTGTCA 3110
Db 240 SerCysGlySerLeuThrProGluArgThrMetSerProIleGlnValLeuAlaValThr 259
QY 3111 ACATCTGCAGTCCCACTGAATCGCTGCTCTCTCTGAATCTCAGGAAGTGCCTATGGGA 3170
Db 260 GlySerAlaSerSerProGluGlnGlyArgSerProGluProThrGluIleSerAlaLys 279
QY 3171 AGGACTATCTCAAGTCACCCCGGAAAAACAACACTGTTCCAGCCCCCGTGGGAAGTAC 3230
Db 280 HisAlaIlePheArgValSerProAspArgGlnSerTrpGlnPheGlnArgSerAsn 299
QY 3231 AACTCCAATGCTAATATCATCACCGGAGACATAATAAATTCACATTCACCTGGGTCT 3290
Db 300 SerAsnSerSerValIleThrThrGluAspAsnLysIleHisIleHisLeuGlySer 319
QY 3291 CAGTTTAAGCGATCTCCTGGCCCTGCCGTGAAGCGGTGAGCCAGTTATCACCGTCCGG 3350
Db 320 ProTyrMetGlnAla-----ValAlaSerProValArgProAlaSerProSerAla 336
QY 3351 CCTGTCAACGTGACAGCGGAGAAGAGGTTTCTACAGGCACAGTCTTCCGTCTCTCCAGG 3410
Db 337 ProLeuGlnAspAsnArgThrGlnGlyLeuIleAsnGlyAlaLeuAsnLysThr----- 354
QY 3411 AACCACTCTCTTCAAGACCCCGGTGCTAGCAAGTGACAGCACTATAACTATAACCCCG 3470
Db 355 -----ThrAsnLysValThrSerSerIleThrIleThrPro 366
QY 3471 GTCACAACGTCTACACACGAGGAACCCAA 3500
Db 367 ThrAlaThrProLeuProArgGlnSerGln 376

RESULT 15
US-10-309-851-22
; Sequence 22, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
; CURRENT APPLICATION NUMBER: US/10/309,851
; CURRENT FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-851-22

Alignment Scores:
Pred. No.: 6.69e-29 Length: 386
Score: 618.50 Matches: 155
Percent Similarity: 54.88% Conservative: 70
Best Local Similarity: 37.80% Mismatches: 126
Query Match: 8.00% Indels: 59
DB: 14 Gaps: 12

US-10-788-793-1 (1-4364) x US-10-309-851-22 (1-386)
QY 2340 GAAGAAGAACTAAGAACAAAGAACATGGGGAGGAGGTCTCAATCTGACCAAGGAGCTA 2399
Db 3 GlnGlnGluAsnArgAsnArgAspLeuGlyArgGluIleGluAsnLeuThrLysGluLeu 22
QY 2400 GAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAACGGCCGAGGATGGTG 2459
Db 23 GluArgTyrArgHisPheSerLysSerLeuArgProSerLeuAsnGlyArgArgIleSer 42
QY 2460 GACGTGCTGTGGCTCCACTGGGTGTCAGACCGAGCGGTGTGTGGGGATGCTGCGGAG 2519

Db 43 AspProGlnValPheSerLysGluValGlnThrGluAlaVal-----Asp 57
QY 2520 GAGGAGACCCCG-----GCTGTGTTTCATTCGCAAA 2549
Db 58 AsnGluProProAspTyrLysSerLeuIleProLeuGluArgAlaValIleAsnGlyGln 77
QY 2550 TCCTTCCAGGAG---GAAAAATCACATCATGAGTAATCTTCGACAGGTAGGCCTGAGAAAA 2606
Db 78 LeuTyrGluGluSerGluAsnGlnAsp-----GluAsp 88
QY 2607 CCCATGGAACGGTCTCTCGGTCTC-----GACAGGTATCCCCCAGCAGCG 2651
Db 89 ProAsnAspGluGlySerValLeuSerPheLysCysSerGlnSerThrProCysProVal 108
QY 2652 AATGAGCTCACCATGAGGAAGTCTTGGATTCTTGGATTCTTGGATTGAGAAAAAGAGAA 2702
Db 109 Asn-----ArgLysLeuTrpIleProTrpMetLysSerLysGluGlyHisLeu 124
QY 2703 ---AACGGTCTCTCCACTCCGAGGAGAAAGGGCCAGGCCAAACAGGGTGCAGGGCAC 2759
Db 125 GlnAsnGly-----LysMetGlnThrLysProAsnAlaAsnPheValGln 139
QY 2760 CCCGGGAGCTGCTCTAGCACCAAGACAGGGCCAGCCCCCTACACATCCGTGTGACACCA 2819
Db 140 ProGlyAspLeuValLeuSerHisThrProGlyGlnProLeuHisIleLysValThrPro 159
QY 2820 GATCATGAGAACAGCACTGCCACCCCTGGAGATCACAAGCCCCACATCTGAA-----GAG 2873
Db 160 AspHisValGlnAsnThrAlaThrLeuGluIleThrSerProThrThrGluSerProHis 179
QY 2874 TTTTCTCTAGTACACCGTCTATCTCTACCTTAGGCAACAGAAACCAAGATAACCATT 2933
Db 180 SerTyrThrSerThrAlaValIleProAsnCysGlyThrProLysGlnArgIleThrIle 199
QY 2934 ATTCATCACCCCAATGTCTATGTCGCAAAAGCCCAAA---AGTGCAGATCTCTACTCTCGGC 2990
Db 200 LeuGlnAsnAlaSerIleThrProValLysSerLysThrSerThrGluAspLeuMetAsn 219
QY 2991 CCAGAACGAGCCATGTCCCTGTCTCAGCATTTACTACTATTTCAGAGAGAGAGCCCGGAA 3050
Db 220 LeuGluGlnGlyMetSerProIleThrMetAlaThrPheAlaArgAlaGlnThrProGlu 239
QY 3051 GGTGAAGGAGCGCCTTTGCCGACAGGCCTGCATCCCCCATCCAAATCATACGGTGTCA 3110
Db 240 SerCysGlySerLeuThrProGluArgThrMetSerProIleGlnValLeuAlaValThr 259
QY 3111 ACATCTGCAGTCCCACTGAAATCGCTGTCTCTCTCTGAATCTCAGGAAGTGCCTATGGGA 3170
Db 260 GlySerAlaSerSerProGluGlnGlyArgSerProGluProThrGluIleSerAlaLys 279
QY 3171 AGGACTATCTCAAAAGTCACCCCGGAAAAACAACACTGTTCCAGCCCCCGTGGGAAGTAC 3230
Db 280 HisAlaIlePheArgValSerProAspArgGlnSerSerTrpGlnPheGlnArgSerAsn 299
QY 3231 AACTCCAATGCTAATATCATCACCGGAGACAAATAAATTCACATTCACCTGGGTCT 3290
Db 300 SerAsnSerSerValIleThrThrGluAspAsnLysIleHisIleHisLeuGlySer 319
QY 3291 CAGTTTAAGCGATCTCTCGGCCTGCCGTGAAGCGGTGAGCCAGTTATCACCGTCCGG 3350
Db 320 ProTyrMetGlnAla-----ValAlaSerProValArgProAlaSerProSerAla 336
QY 3351 CCTGTCAACGTGACAGCGGAGAAGGAGGTTTCTACAGGCACAGTCTTCCGTCTCTCCAGG 3410
Db 337 ProLeuGlnAspAsnArgThrGlnGlyLeuIleAsnGlyAlaLeuAsnLysThr----- 354
QY 3411 AACCACTCTCTTCAAGACCCCGGTGCTAGCAAGTGACAGCACTATAACTATAACCCCG 3470
Db 355 -----ThrAsnLysValThrSerSerIleThrIleThrPro 366
QY 3471 GTCACAACGTCTACACACGAGGAACCCAA 3500

Db 367 ThrAlaThrProLeuProArgGlnSerGln 376

Search completed: September 7, 2004, 15:14:10
Job time : 547.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 7, 2004, 13:29:11 ; Search time 77.5 Seconds
(without alignments)
5814.084 Million cell updates/sec

Title: US-10-788-793-1
Perfect score: 7731
Sequence: 1 ccactgggttttccaaggga.....aaaaaaaaaaaaaaaaaaaa 4364

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10788793/runat_07092004_134022_10839/app_query.fasta_1.4551
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10788793@cgn_1.1.117@runat_07092004_134022_10839 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PTCUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	473.5	6.1	2662 4 US-09-595-684B-31
2	449	5.8	1972 4 US-08-875-435B-3
3	445	5.8	1972 4 US-08-875-435B-4
4	440	5.7	2482 1 US-08-328-254-6
5	436.5	5.6	3248 1 US-08-353-700-1
6	436.5	5.6	3248 5 PCT-US95-16216-1
7	429	5.5	1898 1 US-08-056-200-94
8	429	5.5	1898 2 US-08-800-644-94
9	426	5.5	3878 4 US-09-914-259-11
10	413.5	5.3	1231 4 US-08-714-741-41
11	409.5	5.3	1939 4 US-09-310-187A-1
12	409.5	5.3	2101 1 US-08-466-390-4
Sequence 31, Appli			
Sequence 3, Appli			
Sequence 4, Appli			
Sequence 6, Appli			
Sequence 1, Appli			
Sequence 94, Appli			
Sequence 11, Appli			
Sequence 41, Appli			
Sequence 1, Appli			
Sequence 4, Appli			

13	409.5	5.3	2101	1	US-08-470-950-4	Sequence 4, Appli
14	409.5	5.3	2101	1	US-08-467-781-4	Sequence 4, Appli
15	409.5	5.3	2101	1	US-08-195-487-4	Sequence 4, Appli
16	409.5	5.3	2101	2	US-08-483-924-4	Sequence 4, Appli
17	409.5	5.3	2101	3	US-09-452-294-1	Sequence 1, Appli
18	409.5	5.3	2101	5	PCT-US93-06160-4	Sequence 4, Appli
19	407.5	5.3	1886	4	US-08-938-105-3	Sequence 3, Appli
20	404.5	5.2	2954	4	US-09-150-867-1	Sequence 1, Appli
21	398.5	5.2	2125	4	US-09-919-172-29	Sequence 29, Appli
22	395	5.1	1388	4	US-09-572-191-2	Sequence 2, Appli
23	395	5.1	1388	4	US-09-723-262-2	Sequence 2, Appli
24	395	5.1	1388	4	US-09-723-219-2	Sequence 2, Appli
25	394.5	5.1	1388	4	US-09-976-594-296	Sequence 296, App
26	393	5.1	1695	4	US-09-866-108A-15753	Sequence 15753, A
27	392.5	5.1	1388	2	US-08-685-576-4	Sequence 4, Appli
28	391	5.1	1312	2	US-08-687-080-51	Sequence 51, Appli
29	390.5	5.1	2568	4	US-09-866-108A-3	Sequence 3, Appli
30	390	5.0	1312	2	US-08-592-126-148	Sequence 148, App
31	390	5.0	1312	4	US-09-168-595-148	Sequence 148, App
32	387.5	5.0	1388	2	US-08-685-576-1	Sequence 1, Appli
33	384	5.0	1354	3	US-08-685-871-2	Sequence 2, Appli
34	383	5.0	976	3	US-09-104-324B-4	Sequence 4, Appli
35	378.5	4.9	1581	4	US-09-866-108A-15754	Sequence 15754, A
36	363.5	4.7	1530	4	US-09-976-594-736	Sequence 736, App
37	359	4.6	973	4	US-09-392-714-24	Sequence 24, Appli
38	359	4.6	2468	4	US-09-976-594-726	Sequence 726, App
39	356	4.6	885	2	US-08-533-306A-4	Sequence 4, Appli
40	356	4.6	885	2	US-08-742-923A-4	Sequence 4, Appli
41	350.5	4.5	816	2	US-08-533-306A-6	Sequence 6, Appli
42	350.5	4.5	816	2	US-08-742-923A-6	Sequence 6, Appli
43	350.5	4.5	1093	5	PCT-US93-03077-1	Sequence 1, Appli
44	345.5	4.5	1031	4	US-09-914-259-24	Sequence 24, Appli
45	341	4.4	977	4	US-09-010-147B-18	Sequence 18, Appli

ALIGNMENTS

RESULT 1
US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; TITLE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-31

Alignment Scores:
Pred. No.: 3.19e-24 Length: 2662
Score: 473.50 Matches: 289
Percent Similarity: 39.37% Conservative: 263
Best Local Similarity: 20.61% Mismatches: 461
Query Match: 6.12% Indels: 389
DB: 4 Gaps: 64

US-10-788-793-1 (1-4364) x US-09-595-684B-31 (1-2662)

QY	171	TCAGAA	GATGCA	AAAAA	AGAA	CAAGGCC	CAATCGG	AAGCAGG	AGGAT	---	-----	---	215
		:	:	:									
Db	1381	SerGln	SerLys	GlnGlu	GlnSer	LeuAsn	MetLys	GluLys	AspAsn	GluThr	ThrLys	1400	
		:	:	:									
QY	216	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----		
		:	:	:									
Db	1401	IleVal	SerGlu	MetGlu	GlnPhe	LysPro	LysAsp	SerAla	LeuLeu	ArgIle	GluIle	1420	
		:	:	:									
QY	249	AAACCA	TCTGG	AGAAA	AGTGA	CAAAA	AGACT	AAG	-----	-----	-----	293	
		:		:	:	:	:	:	:	:	:		
Db	1421	GluMet	LeuGly	LeuSer	LysArg	LeuGln	GluSer	HisAsp	GluMet	LysSer	ValAla	1440	
		:		:	:	:	:	:	:	:	:		
QY	294	TTATCC	AAGGAG	GCCTCAT	CCAGCT	CCTG	AGTAT	CATGGA	AGGGAG	TTG	---CAGSCT	350	
		:	:	:	:	:	:	:	:	:	:		
Db	1441	LysGlu	LysAsp	LeuGln	ArgLeu	GlnGlu	ValLeu	SerGln	SerGlu	AspGln	Leu	1460	
		:	:	:	:	:	:	:	:	:	:		
QY	351	CGAGAA	GATGTC	ACATGCT	GAGGAC	AGAG	---AAAA	CCAGCC	CGAGGT	TCTGGAG	407		
		:	:	:	:	:	:	:	:	:	:		
Db	1461	LysGlu	AsnIle	LysGlu	IleVal	AlaLys	HisLeu	GluThr	GluGlu	GluLeu	LysVal	1480	
		:	:	:	:	:	:	:	:	:	:		
QY	408	GCACACT	ATGGATCT	GCAGAACCT	GAGAAA	GTGCTT	CGGGT	CTCTG	CACCGAG	ATGCCATC	467		
		:	:	:	:	:	:	:	:	:	:		
Db	1481	AlaHis	CysCys	LeuLys	GluGln	GluThr	IleAsn	GluLeu	Arg	-----Val	1497		
		:	:	:	:	:	:	:	:	:	:		
QY	468	CTTGCT	CAAGAGA	AGTCC	ATAGGAG	AAGACG	TCTATG	AGAAA	CCCTAT	CTCAGAGCTG	---524		
		:	:	:	:	:	:	:	:	:	:		
Db	1498	AsnLeu	SerGlu	LysGlu	ThrGlu	IleSer	ThrIle	GlnLys	GlnLeu	GluAla	IleAsn	1517	
		:	:	:	:	:	:	:	:	:	:		
QY	525	GACAGACT	GGAGGAAA	AGCAGAA	AGGAGAC	GTACCG	CCGATGCT	PAGAGCAG	CTGCTGCTG	584			
		:	:	:	:	:	:	:	:	:	:		
Db	1518	AspLys	LeuGln	AsnLys	IleGln	GluIle	TrpGlu	LysGlu	---Glu	GlnLeu	AsnIle	1536	
		:	:	:	:	:	:	:	:	:	:		
QY	585	GCTGAG	-----	-----	-----	-----	-----	-----	-----	AAGTGT	CACAGG	602	
		:	:	:	:	:	:	:	:	:	:		
Db	1537	LysGln	IleSer	GluVal	IleGln	GluAsn	ValAsn	GluLeu	LysGln	PheLys	GluHis	Arg1556	
		:	:	:	:	:	:	:	:	:	:		
QY	603	CGC	-----	-----	-----	-----	-----	-----	-----	-----	-----	650	
		:	:	:	:	:	:	:	:	:	:		
Db	1557	LysAla	LysAsp	SerAla	LeuGln	SerIle	GluSer	LysMet	LeuGlu	LeuThr	AsnArg1576		
		:	:	:	:	:	:	:	:	:	:		
QY	651	ATGAACA	AGAGCG	ACGACTT	CACCA	CTGCTG	GAGCAGG	AGCGAG	AGAGGTT	TGAAAAAG	710		
		:	:	:	:	:	:	:	:	:	:		
Db	1577	LeuGln	GluSer	GlnGlu	GluIle	GlnIle	MetIle	LysGlu	LysGlu	GluMet	LysArg1596		
		:	:	:	:	:	:	:	:	:	:		
QY	711	CTCTTGA	ACAAGAAA	AGCTTAC	CAAGCC	CGCAAGAA	-----	AAGGAAA	ACGCTAAG	7			

Db 2052 ArgAspGln-----PheIleAlaThrLeuArgGluMetIleAlaArgAspArgGlnAsn 2069
QY 2127 CACCAAAATGGCCAGCACAAAGCCATA----- 2153
Db 2070 HisGlnValLysProGluLysArgLeuLysArgGlyGlnGlnHisLeuMetGluSer 2089
QY 2154 -----GAGAAA-----GGGAGGCGCTGAGCCAGGAAGCCGAAGTGGGACAC 2195
Db 2090 LeuArgGluLysCysSerArgIleLysGluLeuLysArgTyrSerGluMetAsp 2109
QY 2196 AGGTTTCGGCTGGAGGAGGCTAAAGTCGTGATTACAGGCCGAGGTGCAG----- 2246
Db 2110 HisTyrGluCysLeuAsnArgLeuSerLeuAspLeuGluLysGluIleGluPheHisArg 2129
QY 2247 -----GCTCTCAAGGAGAAGATCCAC 2267
Db 2130 IleMetLysLysLeuLysTyrValLeuSerTyrValThrLysIleLysGluGlnHis 2149
QY 2268 GAGCTGATGAACAAGGAAGACCAGCTGTCTCAGTCCAAGTCGACTATTCGGTCCTTCAG 2327
Db 2150 GluCysIleAsnLys-----PheGlu 2156
QY 2328 CAAAGATTTATGGAAGAAGAAACTAAGAACAAACATGGGAGGAGGTCTCAATCTG 2387
Db 2157 MetAspPheIleAspGluValGluLysGlnLysGluLeuLysIleLysIleGlnHisLeu 2176
QY 2388 ACCAAGGAGCTAGAGCTTTCCAAGCGCTACAGCGGAGCTCTCAGCGCGAGTGGGAACGGC 2447
Db 2177 GlnGlnAspCysAspValProSerArgGluLeuArgAspLeuLysLeuAsnGlnAsn-- 2195
QY 2448 CGAAGGATGGTGACGTGCCTGTGGCCTCCACTGGGGTGCAGACCAGCGGTGTGCGGG 2507
Db 2196 -----MetAspLeuHisIle-----GluGluIleLeuLys 2205
QY 2508 GATGCTCGGAGGAGGAGACCCCGGCTGTGTTCAATTCGAAATCCTTCCAGGAGGAAAT 2567
Db 2206 AspPheSerGluSerGluPheProSer-----IleLysThrGluPheGlnGln----- 2221
QY 2568 CACATCATGAGTAATCTTCGACAGGTA----- 2594
Db 2222 ---ValLeuSerAsnArgLysGluMetThrGlnPheLeuGluGluTrpLeuAsnThrArg 2240
QY 2595 -----GGCCTGAAGAAACCCATGGAACGGTCTCGGTCTCTC 2630
Db 2241 PheAspIleGluLysLeuLysAsnGlyIleGlnLysGluAsnAspArgIleCysGlnVal 2260
QY 2631 -----GACAGGTATCCCCCAGCAGCAATGAG----- 2657
Db 2261 AsnAsnPheAsnAsnArgIleIleAlaIleMetAsnGluSerThrGluPheGluGlu 2280
QY 2658 -----CTCACCATGAGGAAGTCTTGGATTCTTGGATGAGAAAAAGAGAAACGGTCTT 2711
Db 2281 ArgSerAlaThrIleSerLysGluTrp-----GluGlnAspLeu 2293
QY 2712 TCCACTCCGAGGAGAAAGGGCCCAAGCCAGGGTGCAGGGCACCCCGGGGAGCTG 2771
Db 2294 LysSerLeuLysGluLysAsnGluLys----- 2302
QY 2772 GTCCTAGCACCAAGAGGAGGCGCCCTACACATCCGTGTGACACCCAGATCATGAGAAC 2831
Db 2303 -----LeuPheLysAsnTyrGlnThrLeu-----LysThr 2312
QY 2832 AGCACTGCCACCCCTGGAGATCACAAAGCCCACTCTGAAGAGTTTTTCTCTAGTACCACC 2891
Db 2313 SerLeuAlaSerGlyAlaGlnValAsnProThrThrGlnAsp----- 2326
QY 2892 GTCATTCTCTACCTAGGCAACCAAGAACCAAGAATAACCATATTTCATCACCCTCAATGTC 2951
Db 2327 -----AsnLysAsnProHisValThr---SerArgAlaThrGlnLeu 2339
QY 2952 ATGTCGCAAAAGCCCAAAAGTGCAGATCCTTACTCTC-----GGCCAGAACGAGCCATG 3005
Db 2340 ThrThrGluLysIleArgGluLeuGluAsnSerLeuHisGluAlaLysGluSerAlaMet 2359

QY 3006 TCC-----CCTGTCACGATTACTACTATT 3029
Db 2360 HisLysGluSerLysIleIleLysMetGlnLysGluLeuGluValThrAsnAspIleIle 2379
QY 3030 TCCAGAGAGAAGACCCCGGAAGGTGGAAGGAGCGCCTTTGCCGACAGGCCCTGCATCCCC 3089
Db 2380 AlaLysLeuGlnAlaLysValHisGluSerAsnLysCysLeuGluLysThrLysGluThr 2399
QY 3090 ATCCAAATCATGACGGTGTCAACATCTGCAGCTCCCACCTGAAATCGCTGTCTCTCTGAA 3149
Db 2400 IleGlnValLeuGlnAspLysValAla-----LeuGlyAlaLysProTyr 2414
QY 3150 TCTCAGGAAGTG---CCTATGGGAAGGACTATCTCAAAAGTCACCCCGGAAAAACAAC 3206
Db 2415 LysGluGluIleGluAspLeuLysMetLysLeuValLysIleAspLeuGluLysMetLys 2434
QY 3207 GTTCCAGCCCCCGTCGGAAGTACAACCTCCAATGCTAATATCATCACCACGGAAGACAAT 3266
Db 2435 AsnAlaLysGluPheGluLys---GluIleSerAlaThrLysAlaThrValGluTyrGln 2453
QY 3267 AAAATTTCACATTACCTG-----GGTTCTCAGTTTAAAGCGATCTCCT 3308
Db 2454 LysGluValIleArgLeuLeuArgGluAsnLeuArgArgSerGlnGlnAlaGlnAspThr 2473
QY 3309 GGGCCTGCCGTGAAGCGGTGAGCCCAAGTTATCACCCTCCGGCCTGTC----- 3356
Db 2474 SerValIleSerGluHisThrAspProGlnProSerAsnLysProLeuThrCysGlyGly 2493
QY 3357 -----AACGTGACAGCG 3368
Db 2494 GlySerGlyIleValGlnAsnThrLysAlaLeuIleLeuLysSerGluHisIleArgLeu 2513
QY 3369 GAGAAGGAGGTTTCT-----ACAGGCACAGTC 3395
Db 2514 GluLysGluIleSerLysLeuLysGlnGlnAsnGluGlnLeuIleLysGlnLysAsnGlu 2533
QY 3396 CTTCGCTCTCCAGGAACCACTCTCTTCAAGA----- 3428
Db 2534 LeuLeuSerAsnAsnGlnHisLeuSerAsnGluValLysThrTrpLysGluArgThrLeu 2553
QY 3429 -----CCCGGTGCTAGCAAAAGTGACC 3449
Db 2554 LysArgGluAlaHisLysGlnValThrCysGluAsnSerProLysSerProLysValThr 2573
QY 3450 AGCACTATAACT-----ATAACCCCGGTCAAA-----CGTCAT----- 3483
Db 2574 GlyThrAlaSerLysLysLysGlnIleThrPro-SerGlnCysLysGluArgAsnLeuG1 2593
QY 3484 -----CCACACGAGGAACCCCAATCAGTGT---CAGGACAAGATGGTCTCTCAG 3530
Db 2593 nAspProValProLysGluSerProLysSerCysPheAspSerArgSerLys-SerL 2613
QY 3531 CGGCCTACCCCAACCCGCTTCTCTATGTCAAAAGGTATGAA----- 3571
Db 2613 euProSerProHisProValArgTyrPheAspAsnSerSerLeuGlyLeuCysProGluV 2633
QY 3572 -----AGCTGGAAGCCAGTAGTG 3590
Db 2633 alGlnAsnAlaGlyAlaGluSerValAspSerGlnProGlyProThrHisAlaSerSerG 2653
QY 3591 GC 3592
Db 2653 1y 2653

RESULT 2
US-08-875-435B-3
; Sequence 3, Application US/08875435B
; Patent No. 6593304
; GENERAL INFORMATION:
; APPLICANT: Hasegawa, Kazuhide
; APPLICANT: Arakawa, Emi
; APPLICANT: Oda, Shoji

```
; APPLICANT: Matsuda, Yuzuru
; APPLICANT: Takahashi, Katsuhito
; APPLICANT: Sugahara, Michihiro
; APPLICANT: Ishiyama, Haruo
; TITLE OF INVENTION: RECOMBINANT DNA COMPRISING DNA CODING
; TITLE OF INVENTION: FOR MYOSIN HEAVY CHAIN SM1 ISOFORM PROTEIN INSERTED INTO
; TITLE OF INVENTION: VECTOR DNA, MICROORGANISM CARRYING THE RECOMBINANT DNA, AND
; TITLE OF INVENTION: AN AGENT FOR TREATMENT OF ARTERIOSCLEROSIS COMPRISING THE
; TITLE OF INVENTION: RECOMBINANT DNA
; FILE REFERENCE: 07898-013001
; CURRENT APPLICATION NUMBER: US/08/875.435B
; CURRENT FILING DATE: 1997-07-25
; PRIOR APPLICATION NUMBER: PCT/JP96/00134
; PRIOR FILING DATE: 1996-01-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Mus musculus
US-08-875-435B-3

Alignment Scores:
Pred. No.:      1.37e-22      Length:      1972
Score:          449.00      Matches:      203
Percent Similarity: 42.82%      Conservative: 167
Best Local Similarity: 23.50%      Mismatches:  292
Query Match:      5.81%      Indels:      202
DB:               4          Gaps:         33

US-10-788-793-1 (1-4364) x US-08-875-435B-3 (1-1972)
QY 276 ACTAAGAAGTCTGTGAGTTATCCAAGGAGGACCTCATCCAG----- 317
Db 807 ThrLysArgGlnGlnGlnLeuThrAlaMetLysValilleGlnArgAsnCysAlaAlaTyr 826
QY 318 -----CTCCTGAGT 326
Db 827 LeuLysLeuArgAsnTrpGlnTrpTrpArgLeuPheThrLysValLysProLeuLeuGln 846
QY 327 ATC-----ATGGAAGGGAGTTGCAGGCTCGAGAAGATGTCATCCACATGCTG----- 374
Db 847 ValThrArgGlnGluGluMetGlnAlaLysGluGluMetGlnLysIleThrGlu 866
QY 375 AGGACACAGAAAAACCAGCCCGAGGTTCTGGAGGCACACTATGGATCTGCAGAACCTGAG 434
Db 867 ArgGlnGlnLysAlaGluThrGluLeuLysGluLeuGlu----- 879
QY 435 AAAGTGCTTCGGGTCTGCACCAGATGCCATCCTTGCTCAAGAGAAGTCCATAGGAGAA 494
Db 880 -----GlnLysHisThrGlnLeuAlaGluGluLysThrLeu----- 891
QY 495 GACGTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAAAAAGCAGAGACG 554
Db 892 ---LeuGlnGluGlnLeuGlnAlaGluThrGluLeuTyrAlaGluSerGluGluMetArg 910
QY 555 TACCGCGGCATGCTAGACAGCTGCTGCTGGCTGAGAAAGTGTACAGGCGCACCGTGTAC 614
Db 911 ValArg-----LeuAlaAlaLysLys-----Gln 918
QY 615 GAGCTGGAGAACGAGAGCAACAGCACACTGACTACATGAACAAGACGACGACTTCACC 674
Db 919 GluLeuGluGluIleLeuHisGluMetGluAlaArgLeuGluGluGluAspArgArg 938
QY 675 AACCTGCTGGAGCAGGACGAGAGG-----TTGAAAAAGCTC 713
Db 939 GlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGlnMetLeuAspLeuGluGln 958
QY 714 CTTGAACAAGAAAAAGCTTACCAAGCCGCAAGAAAAAGGAAAAACGCT-----AAG 764
Db 959 LeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeuLysValThrAlaGluAla 978
QY 765 CGGCTCAACAAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCATGTTGTG 824
```

```
979 LysIleLysLysLeuGluAspAspIle-----LeuValMetAsp 991
825 GACGAGAGGCAGATGCACATCGAGCAACTGGCCTGCAGAGTCAGAAAAGTCCAGGACCTC 884
992 AspGlnAsnSerLysLeuSerLysGluArgLysLeuLeuGluGluArgValSerAspLeu 1011
885 ACTCAAAAGCTGAGGGAGGAGGAAGAAAAACTCAAACGGGTCACT----- 929
1012 ThrThrAsnLeuAlaGluGluGluLysAlaLysAsnLeuThrLysLeuLysSerLys 1031
930 -----TACAAATCCAAAGGAAGACCGCCAG 953
1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051
954 AAGCTGCTCAAGTTAGAAGTGGACTTCGAACACACAAGGCTCGAGGTTTCCAGGAGCAC 1013
1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071
1014 GAAGAGATGAACGCCAAATTTGGCGAATCAAGAATCTCACAACCCGGCAACTTCGACTCAA 1073
1072 AlaAspLeuGlnAlaGlnIleAla-----GluLeuLysMetGln 1084
1074 CTG-----GTTGGCTTATCGCAAAAGGATTGAGGAGCTGGAA 1109
1085 LeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgLeuAspGluGluIleAla 1104
1110 GAGACCAATAAAAGCCTTCAGAAG-----CTTGAGATGGAGGGCAAG 1139
1105 GlnLysAsnAsnAlaLeuLysLysIleArgGluLeuGluAspThrLeuAspSerThrAlaThr 1124
1140 GAAGAGCTCCAGGAGCTGAGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAAACTCCAGTCTC 1199
1125 GluAspLeuAspSerGluArgAlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeu 1144
1200 ATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTG-----CTTGAGATGGAGGGCAAG 1250
1145 GlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeuAspSerThrAlaThr 1164
1251 GATGAAGAGATC---ACGAAGACCCGAGGCCCATGTCGGGAGCTGAAGAAGAAGCTCCAA 1307
1165 GlnGlnGluLeuArgAlaLysArgGluGlnGluValThrValLeuLysLysAlaLeuAsp 1184
1308 GAGGAA---GAACACCACAGCAAGGAACTTAGACTAGAGTGGAGAGCTGCAGAAGAGG 1364
1185 GluGluThrArgSerHisGluAlaGlnValGlnGluMetArgGlnLysHisThrGlnAla 1204
1365 ATGTCTGAGCTGGAGAAAGCTGGAGGAAGCGTTTCAGCCGAGTAAGTCGGAATGCACCCAG 1424
1205 ValGluGluLeuThrGluGlnLeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLys 1224
1425 CTCCATCTGAACCTGGAGAAGGAGAGAACCTTAACCAAGACCTGCTGAACGAGCTGGAG 1484
1225 SerLysGlnThrLeuGluLysGlu-----AsnAlaAspLeuAlaGlyGluLeuArg 1241
1485 GTG-----GTCAAGAGTCGAGTTAAAGAACTCGAA----- 1514
1242 ValLeuGlyGlnAlaLysGlnGluValGluHisLysLysLysLysLeuGluValGlnLeu 1261
1515 -----TGTCGCGAGAGTAGACTGGAGAAGGCCGAGTTAAGCCTCAA 1556
1262 GlnAspLeuGlnSerLysCysSerAspGlyGluArgAlaArgAlaGluLeuSer----- 1279
1557 GATGACCTTACAAAGCTGAAGTCCCTTCACTGTGATGCTGGTGGATGAGAGGAAAAATATG 1616
1280 AspLysValHisLysLeuGln-----AsnGluValGluSerVal 1292
1617 ATGGAGAAAAATAAGCAAGAAGAGAGGAAAGTGGATGGTGTGAATAAAAACTTTAAGGTG 1676
1293 ThrGlyMetLeuAsnGluAlaGluGlyLysAlaIleLysLeuAlaLysAspValAlaSer 1312
1677 GAGCAGGGAAAAAGTCATGGATGTGACGGAAAAAGCTAATCGAGGAAAGCAAGAGCTTTTA 1736
```

Db 1313 LeuGlySerGlnLeuGlnAspThrGlnGluLeuLeuGlnGluThrArgGlnLysLeu 1332
QY 1737 AAACCTCAATCT-----GAAATGGAGGAGGAGGAGTACAGTCTGACAAAAGGAGAGG 1787
Db 1333 AsnValSerThrLysLeuArgGlnLeuGluAspGluArgAsnSerLeuGlnAspGlnLeu 1352
QY 1788 GATGAGCTGATGGGTAAACTGAGGAGCGGAAGAAGAGGCTCTGTGAACCTGAGCTGCAGT 1847
Db 1353 AspGluGluMetGluAlaLysGlnAsnLeuGluArgHisValSerThrLeuAsnIleGln 1372
QY 1848 GTAGACTTACTAAAGAAGCGGCTT-----GAT 1874
Db 1373 LeuSerAspSerLysLysLeuGlnAspPheAlaSerThrIleGluValMetGluGlu 1392
QY 1875 GGCATAGAGGAGGTAGAAAGGGAATAAACCAGGAGGTAGGTCTGTCAAGGGTCTGAGTTC 1934
Db 1393 GlyLysLysArgLeuGlnLysGluMet--GluGlyLeuSerGlnGlnTyrGluGluLys 1411
QY 1935 ACCTGCCCGGAAGAC-----AATAAGATC----- 1958
Db 1412 AlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAspAsp 1431
QY 1959 -----AGAGAACTAACGCTTGAATCGAGAGACTGAAGAAA 1994
Db 1432 LeuValValAspLeuAspAsnGlnArgGlnLeuValSerAsnLeuGluLysLysGlnLys 1451
QY 1995 CGGCTCCAGCAGTTGGAGGTGTGGAGGGGACTTG---ATGAAGACCGAGGACGAATAT 2051
Db 1452 LysPheAspGlnLeuLeuAlaGluGluLysAsnIleSerSerLysTyrAlaAspGluArg 1471
QY 2052 GACCAGTTGGAGCAGAAAGTTCAAGACCGAGCAGGATAAGGCAAACTTCTCTCCCGCAG 2111
Db 1472 AspArgAlaGluAlaGluAlaArgGluLysGluThrLysAlaLeuSerLeuAlaArgAla 1491
QY 2112 CTCGAG-----GAATCAAAACACCAATATGCCCAAGCACAAAGCCATAGAGAAAGGG 2162
Db 1492 LeuGluGluAlaLeuGluAlaLysGluGluLeuGluArgThrAsnLysMetLeuLysAla 1511
QY 2163 GAG-----GCCGTGAGCCAGGAAGCCGAACCTGCGACACAGGTTTCGGCTGGAGGAG 2213
Db 1512 GluMetGluAspLeuValSerSerLysAspAspValGlyLysAsnValHisGluLeuGlu 1531
QY 2214 GCTAAAAGTCGTGATTTACAGCCGAGGTGCAGGCTCTCAAGGAGAAAGATCCACGAG--- 2270
Db 1532 LysSerLysArgAlaLeuGluThrGlnMetGluGluMetLysThrGlnLeuGluGluSer 1551
QY 2271 -----CTGATGAACAAGGAAGACCGAGCTGTCTCAGCTCCAAGTCGACTATTCCGTC 2321
Db 1552 GluAspAspValGlnAlaThrGluAspAlaLysLeuArgLeuGluValAsnMetGlnAla 1571
QY 2322 -----CTTCAGCAAAAGATTATGGAAGAAGAAACTTAAGAAC 2357
Db 1572 LeuLysGlyGlnPheGluArgAspLeuGlnAlaArgAspGluGlnAsnGluGluLysArg 1591
QY 2358 AAGAACATGGGAGGAGGTCTCAATCTGACCAAGGAGCTAGAGCTTTCCAAGCGCTAC 2417
Db 1592 ArgGlnLeuGlnArgGlnLeuHisGluTyrGluThrGluLeuGlu---AspGluArgLys 1610
QY 2418 AGCCGAGCTCTC 2429
Db 1611 GlnArgAlaLeu 1614

RESULT 3

US-08-875-435B-4
; Sequence 4, Application US/08875435B
; Patent No. 6593304
; GENERAL INFORMATION:
; APPLICANT: Hasegawa, Kazuhide
; APPLICANT: Arakawa, Emi
; APPLICANT: Oda, Shoji
; APPLICANT: Matsuda, Yuzuru
; APPLICANT: Takahashi, Katsuhito
; APPLICANT: Sugahara, Michihiro

; APPLICANT: Ishiyama, Haruo
; TITLE OF INVENTION: RECOMBINANT DNA COMPRISING DNA CODING
; TITLE OF INVENTION: FOR MYOSIN HEAVY CHAIN SMI ISOFORM PROTEIN INSERTED INTO
; TITLE OF INVENTION: VECTOR DNA, MICROORGANISM CARRYING THE RECOMBINANT DNA, AND
; TITLE OF INVENTION: AN AGENT FOR TREATMENT OF ARTERIOSCLEROSIS COMPRISING THE
; TITLE OF INVENTION: RECOMBINANT DNA
; FILE REFERENCE: 07898-013001
; CURRENT APPLICATION NUMBER: US/08/875,435B
; CURRENT FILING DATE: 1997-07-25
; PRIOR APPLICATION NUMBER: PCT/JP96/00134
; PRIOR FILING DATE: 1996-01-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-08-875-435B-4

Alignment Scores:

Pred. No.: 2.58e-22 Length: 1972
Score: 445.00 Matches: 248
Percent Similarity: 35.79% Conservative: 164
Best Local Similarity: 21.55% Mismatches: 333
Query Match: 5.76% Indels: 406
DB: 4 Gaps: 44

US-10-788-793-1 (1-4364) x US-08-875-435B-4 (1-1972)

QY 192 AAGCCCAATCGGAAGGAGGAGGATGTATGGCTTCC-----GGAACTATCAAA 239
Db 846 GlnValThrArgGlnGluGluMetGlnAlaLysGluAspGluLeuGlnLysIleLys 865
QY 240 ---AGGCACCTCAAAACCATCTGGAGAAAGTGAGAAAAAAGACTAAGAAGTCTGTGGAGTTA 296
Db 866 GluArgGlnGlnLysAlaGluSerGluLeuGlnGluLeuGlnLysHisThrGlnLeu 885
QY 297 TCC----- 299
Db 886 SerGluGluLysAsnLeuLeuGlnGluGlnAlaGluThrGluLeuTyrAlaGlu 905
QY 300 -----AAGGAGGACCTCATCCAGCTCCTGAGT 326
Db 906 AlaGluGluMetArgValArgLeuAlaAlaLysLysGlnGluLeuGluGluIleLeuHis 925
QY 327 ATCATGGAAGGGAGTTGTCAGGCTCGAGAAGATGTCTCATCCACATGTGAGGACAGAGAA 386
Db 926 GluMetGluAlaArgLeuGluGluGluAspArgGlyGlnGlnLeuGlnAlaGluArg 945
QY 387 ACCAAG-----CCGAGGTTCTGGAGGCACACTATGGATCTGCAGAACTGAGAAAGTG 440
Db 946 LysLysMetAlaGlnGlnMetLeuAspLeuGluGlnLeuGluGluGluAlaAla 965
QY 441 CTTCCGGTCTCCACCGAGATGCCATCCTTGTCTCAAGAG-----AAGTCCATAGAGAA 494
Db 966 ArgGlnLysLeuGlnLeuGluLysValThrAlaGluAlaLysIleLysLysLeuGluAsp 985
QY 495 GACGTCTATGAGAAACCTATCTCAGAGCTGGACAGACTGGAGGAAAGCAGACGAGACG 554
Db 986 AspIleLeuValMetAsp-----AspGlnAsnAsnLysLeuSerLysGlu--- 1000
QY 555 TACCGCCGCATGCTAGAGCAGCTGCTG-----CTGGCTGAGAAAGTG 596
Db 1001 ---ArgLysLeuLeuGluGluArgIleSerAspLeuThrThrAsnLeuAlaGlu----- 1017
QY 597 CACAGGCGCACCGTGTACGAGCTGGAGAACGAGAAAGCACAAGCAC---ACTGACTACATG 653
Db 1018 -----GluGluGluLysAlaLysAsnLeuThrLysLeuLys 1029
QY 654 AACAGAGCGACGACTTCACCAACCTGCTGGAG----- 686
Db 1030 AsnLysHisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSer 1049

QY 687 ---CAGGAGCGAGAGGTTGAAAAAGCTCCTTGAACAGAA----- 725
Db 1050 ArgGlnGluLeuGluLysLeuLysArgLysMetAspGlyGluAlaSerAspLeuHisGlu 1069
QY 726 -----AAAGCTTACCAGCCCGCAAGAA 749
Db 1070 GlnIleAlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGlu 1089
QY 750 AAGGAAACGCTAAGCGGCTCAACAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCTTC 809
Db 1090 GluGluLeuGlnAlaAlaLeuAlaArgLeuGluAspGlu---ThrSerGlnLysAsnAsn 1108
QY 810 GCCCTCATGTTGGTGGACGAGAGGCAGATGCACATCGAGCAACTG-----GGCCTGCAG 863
Db 1109 AlaLeuLysLysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAsp 1128
QY 864 AGTCAG-----AAAGTCCAGGACCTCACTCAGAAAGCTGAGGAGGAGGAA 908
Db 1129 SerGluArgAlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeu 1148
QY 909 GAAAAACTCAAAGCG-----GTCACTTACAATCCAAGGAAGAC 947
Db 1149 GluAlaLeuLysThrGluLeuGluAspThrLeuAspThrThrAlaThrGlnGlnGluLeu 1168
QY 948 CGCCAGAG-----CTGCTCAAGTTAGAAGTGGACTTCGAA----- 983
Db 1169 ArgAlaLysArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArg 1188
QY 984 ---CACAAAGGCC-----TCGAGGTTTTCCAGGAGCAGCAAGAGATGAACGCCAAATTG 1034
Db 1189 SerHisGluAlaGlnValGlnGluMetArgGlnLysHisThrGlnValGluGluLeu 1208
QY 1035 GCGAATCAA----- 1043
Db 1209 ThrGluGlnLeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysThrLysGlnThr 1228
QY 1044 ---GAATCTCACAAACCGCAACTTCGACTCAAACTGGTTGGCTTATCGCAAAGGATTGAG 1100
Db 1229 LeuGluLysGluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGln 1248
QY 1101 GAGCTGGAAGAGACCATAAAAGCCTTCAGAAAGCCAGAGGAAGAGCTCCAGGAGCTGAGA 1160
Db 1249 GluValGluHisLysLysLys-----LysLeuGluValGlnLeuGlnGluLeuGln 1265
QY 1161 GAGAAAATGCCAAAGGGGAATGTGGAACTCC-----AGTCTC 1199
Db 1266 SerLysCysSerAspGlyGluArgAlaArgAlaGluLeuAsnAspLysValHisLysLeu 1285
QY 1200 ATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTCTTGAGATGGAGGGCAAG----- 1250
Db 1286 GlnAsnGluValGluSerValThrGlyMetLeuSerGluAlaGluGlyLysAlaIleLys 1305
QY 1251 ---GATGAAGAGATCACGAAGACCGAGGCCAGTCCGGGAGCTGAAGAGAAGCTCCAA 1307
Db 1306 LeuAlaLysGluValAlaSerLeuGlySerGlnLeuGlnAspThrGlnGluLeuGln 1325
QY 1308 GAGGAA----- 1313
Db 1326 GluGluThrArgGlnLysLeuAsnValSerThrLysLeuArgGlnLeuGluAspGluArg 1345
QY 1314 -----GAACACCCAC 1322
Db 1346 AsnSerLeuGlnGluGlnLeuAspGluMetGluAlaLysGlnAsnLeuGluArgHis 1365
QY 1323 AGCAAGGAACCTTAGACTAGAAGTG-----GAGAAGCTGCAGAAGAGGATGTCT 1370
Db 1366 IleSerThrLeuAsnIleGlnLeuSerAspSerLysLysLysLeuGlnAspPheAlaSer 1385
QY 1371 GAGCTGGGAAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTC--- 1427
Db 1386 ThrValGluSerLeuGluGlyLysLysArgPheGlnLysGluIleGluSerLeuThr 1405
QY 1428 -----CATCTGAACCTGGAGAAGGAGAAACCTAAC 1460

Db 1406 GlnGlnTyrgluGluLysAlaAlaIaTyrgluLysLeuGluLysThrLysAsnArgLeu 1425
QY 1461 AAAGACTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCC 1520
Db 1426 GlnGlnGluLeuAspAspLeuValValAspLeuAspAsnGlnArgGlnLeu----- 1442
QY 1521 GAGAGTAGACTGGAGAAGCCGAGTTAAGCCTCAAAAGATGACCTTACAAAAGCTGAAAGTCC 1580
Db 1443 ValSerAsnLeuGluLysLysGlnLysLysPheAspGlnLeuLeuAlaGluLysAsn 1462
QY 1581 TTCACTGTGATGCTGTGTGATGAGAGGAAAAATATGTGGAGAAAAATAAAGCAAGAAGAG 1640
Db 1463 IleSerSerLysTyrgluAlaAspGluArgAspArgAlaGluAlaGluAlaArgGluLysGlu 1482
QY 1641 AGGAAA-----GTG 1649
Db 1483 ThrLysAlaLeuSerLeuAlaArgAlaLeuGluGluAlaLeuGluAlaLysGluGluLeu 1502
QY 1650 GATGGGTTGAATAAAAACTTTAAGGTGGAGCAGGGGAAAAGTCATG-----GAT 1697
Db 1503 GluArgThrAsnLysMetLeuLysAlaGluMetGluAspLeuValSerSerLysAspAsp 1522
QY 1698 GTGACGGAAAAGCTAATCGAGGAAAGCAAGAGCTTTTAAACTCAAATCTGAAATGGAG 1757
Db 1523 ValGlyLysAsnValHisGluLeuGluLysSerLysArgAlaLeuGluThrGlnMetGlu 1542
QY 1758 GAAAAGGAGTACAGTCTGACAAAAGGAGGAGGATGAGCTGATG-----GGTAAA 1805
Db 1543 GluMetLysThrGlnLeuGluGluLeuGluAspGluLeuGlnAlaThrGluAspAlaLys 1562
QY 1806 CTGAGGAGCGAA----- 1817
Db 1563 LeuArgLeuGluValAsnMetGlnAlaLeuLysValGlnPheGluArgAspLeuGlnAla 1582
QY 1818 -----GAAGAAAGGTCTGTGAACCTGAGCTGCAGTGTAGAC----- 1853
Db 1583 ArgAspGluGlnAsnGluGluLysArgArgGlnLeuGlnArgGlnLeuHisGluTyrglu 1602
QY 1854 -----TTACTAAAGAAGCGGCTT 1871
Db 1603 ThrGluLeuGluAspGluArgLysGlnArgAlaLeuAlaAlaAlaLysLysLysLeu 1622
QY 1872 GATGGC----- 1877
Db 1623 GluGlyAspLeuLysAspLeuGluLeuGlnAlaAspSerAlaIleLysGlyArgGluGlu 1642
QY 1878 -----ATAGAGGAGGTAGAAAAGGAAATAAAC 1904
Db 1643 AlaIleLysGlnLeuLeuLysLeuGlnAlaGlnMetLysAspPheGlnArgGluLeuGlu 1662
QY 1905 CGAGGTAGGTCTGTCAGGGGTCTGAGTTCACCTGCCCGAAGACAATAAGATCAGAGAA 1964
Db 1663 AspAlaArgAlaSerArgAspGluIlePheAlaThrAlaLysGluAsnGluLysLysAla 1682
QY 1965 CTAACGCTTGAA-----ATCGAGAGACTG 1988
Db 1683 LysSerLeuGluAlaAspLeuMetGlnLeuGlnGluAspLeuAlaAlaGluArgAla 1702
QY 1989 AAGAAACGG-----CTCCAGCAGTTGGAGGTGGTGAG----- 2021
Db 1703 ArgLysGlnAlaAspLeuGluLysGluGluLeuAlaGluLeuAlaSerSerLeuSer 1722
QY 2022 -----GGGGACTTGATGAAGACCGAGCAGCAATATGACCAGTTGGAG 2063
Db 1723 GlyArgAsnAlaLeuGlnAspGluLysArgArgLeuGluAlaArgIleAlaGlnLeuGlu 1742
QY 2064 CAGAAAGTTCAGAACCGAGCAG-----GATAAGGCCAAACTTCCTC 2102
Db 1743 GluGluLeuGluGluGlnGlyAsnMetGluAlaMetSerAspArgValArgLysAla 1762
QY 2103 TCCCAGCAGCTCGAGGAAATCAACACCAAATGGCCAAAGCACAAAGCC---ATAGAGAAA 2159
Db 2159

Db 1763 ThrGlnGlnAlaGluGlnLeuSerAsnGluLeuAlaThrGluArgSerThrAlaGlnLys 1782
QY 2160 GGGAGGCCGTGAGCCAGGAAGCCGAA----- 2186
Db 1783 AsnGluSerAlaArgGlnGlnLeuGluArgGlnAsnLysGluLeuLysSerLysLeuGln 1802
QY 2187 -----CTGGACACAGGTTTCGG----- 2204
Db 1803 GluMetGluGlyAlaValLysSerLysPheLysSerThrIleAlaAlaLeuGluAlaLys 1822
QY 2205 -----CTGGAG-----GAGGCTAAAGTCGTGATTACAGGCCGAGGTGCAG 2246
Db 1823 IleAlaGlnLeuGluGlnValGluGlnAlaArgGluLysGlnAlaAlaLys 1842
QY 2247 GCTCTCAAGGAG-----AAGATCCACGAGCTGATG----- 2276
Db 1843 AlaLeuLysGlnArgAspLysLysLeuLysGluMetLeuLeuGlnValGluAspGluArg 1862
QY 2276 ----- 2276
Db 1863 LysMetAlaGluGlnTyrLysGluGlnAlaGluLysGlyAsnAlaLysValLysGlnLeu 1882
QY 2277 -----AACAAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTATTCGGTC 2321
Db 1883 LysArgGlnLeuGluGluAlaGluGluSerGlnArgIleAsnAlaAsnArgArgLys 1902
QY 2322 CTTCAAGAAAGATTTATGGAAGAAGAACTAAGAACAAAGAACATGGGAGGGAGGTCTCTC 2381
Db 1903 LeuGlnArgGluLeuAspGluAlaThrGluSerAsnGluAlaMetGlyArgGluValAsn 1922
QY 2382 AATCTGACCAAGGAGCTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGG 2441
Db 1923 AlaLeuLysSerLysLeuArgArgGlyAsnGluThrSerPheValProThrArgArgSer 1942
QY 2442 AACGGCCGAAGGATGGTGGACGTGCCTGTGGCCTCCACTGGGGTGCAGACCGAGCGGTG 2501
Db 1943 GlyGlyArgArgValIleGlu----- 1949
QY 2502 TCGGGGATGCTCGGAGGAGGAGACCCCGGCT 2534
Db 1950 AsnAlaAspGlySerGluGluGluValAspAla 1960

RESULT 4

US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815

; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-328-254-6
Alignment Scores:
Pred. No.: 6.14e-22 Length: 2482
Score: 440.00 Matches: 306
Percent Similarity: 35.59% Conservative: 245
Best Local Similarity: 19.77% Mismatches: 521
Query Match: 5.69% Indels: 476
DB: 1 Gaps: 61
US-10-788-793-1 (1-4364) x US-08-328-254-6 (1-2482)
QY 51 TTAAGGAGTCGACAAACAGGTGGGAATGAGATCACGAAAT-----CAAGGTGGAGAAAGT 104
Db 1018 LeuSerSerArgSerLeuLeuGlyIleAspThrGluAspAlaIleGlnGlyArgAsnGlu 1037
QY 105 TCA-----TCTAACGGGCATGTCTCTCTGCCCCAAGTCTCTCCATCATCAGCAGTGAT 155
Db 1038 SerCysAspIleSerLysGluHisThrSerGluThrThrGluArgThrProLysHisAsp 1057
QY 156 GGTGTAAGGGCCCTCAGAAGATGCAAAAAGAACAAAGCCCAATCGGAAGGAGGAGGAT 215
Db 1058 ValHisGlnIleCysAspLysAspAlaGlnGlnAsp---LeuAsnLeuAspIleGluLys 1076
QY 216 GTCATGGTTCGGAACCTATCAAAAGGCACCTCAAAACCATCTGGAGAA-----AGTGAG 269
Db 1077 IleThrGluThrGlyAla-----LeuLysProThrGlyGluCysSerGlyGlu 1092
QY 270 AAAAAGACTAAGAAAGTCTGTGGAGTTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATC 329
Db 1093 GlnSerProAspThrAsnTyrGluProProGlyGluAspLysThrGlnGlySerSerGlu 1112
QY 330 ATGGAAGGGAGTTGCAG-----GCT 350
Db 1113 CysIleSerGluLeuSerPheSerGlyProAsnAlaLeuValProMetAspPheLeuGly 1132
QY 351 CGAGAAGATGTCATCCAC-----ATGCTGAGGACAGAGAAAACCAAGCCCGAGTTCTG 404
Db 1133 AsnGlnGluAspIleHisAsnLeuGlnLeuArgValLysGluThrSerAsnGluAsnLeu 1152
QY 405 GAG---GCACACTATGGATCTGCAGAACCTCAGAAAAGTGTTCGGGTCTCGACCGA--- 458
Db 1153 ArgLeuLeuHisValIleGluAspArgAspArgLysValGluSerLeuLeuAsnGluMet 1172
QY 459 -----GATGCC 464
Db 1173 LysGluLeuAspSerLysLeuHisLeuGlnGluValGlnLeuMetThrLysIleGluAla 1192
QY 465 ATCCTTGCTCAAGAGAAAGTCCATAGGAGAGACGCTCTATGAGAAACCTATCTCAGAGCTG 524
Db 1193 CysIleGluLeuGluLysIleValGlyGlu-----LeuLysLysGluAsnSerAspLeu 1210
QY 525 GACAGACTGGAGAAAAGCAGAAAGGAGACGTACCGCCGATGCTAGACGAGCTGCTGCTG 584
Db 1211 SerGluLysLeuGluTyrPheSerCysAspHisGlnGluLeuLeuGlnArgValGluThr 1230
QY 585 GCTGAGAAGTGTACAGGGCCACCGTGTACGAGCTGGAGAACGAGAACGACACACACT 644
Db 1231 SerGlu-----GlyLeuAsnSerAspLeuGluMetHisAla 1242
QY 645 GACTACATGAACAAGAGCGACGACTTCACCAACCTG----- 680
Db 1243 AspLysSerSerArgGluAspIleGlyAspAsnValAlaLysValAsnAspSerTrpLys 1262

Qy	681	-----CTGGAGCAGGAGCGAGAGGTTG-----AAAAAGTCCTT	716
Db	1263	GluArgPheLeuAspValGluAsnGluLeuSerArgIleArgSerGluLysAlaSerIle	1282
Qy	717	GAACAAGAAAAAGCTTACCAA-----	737
Db	1283	GluHisGluAlaLeuTyrLeuGluAlaAspLeuGluValGlnThrGluLysLeuCys	1302
Qy	738	GCCCGCAAGAAAAAGCAACGCTAAGCGGCTC-----	770
Db	1303	LeuGluLysAspAsnGlnLysGlnLysValIleValCysLeuGluGluLeuSer	1322
Qy	771	-----AACAAACTTCGAGATGAGCTT-----GTGAAGCTCAAGTCC	806
Db	1323	ValValThrSerGluArgAsnGlnLeuArgGlyGluLeuAspThrMetSerLysThr	1342
Qy	807	TCGCGCTCATGTTGGTGGACGAGAGGCAG-----	836
Db	1343	ThrAlaLeuAspGlnLeuSerGluLysMetLysGluLysThrGlnGluLeuGluSerHis	1362
Qy	837	-----ATGCAC-----	848
Db	1363	GlnSerGluCysLeuHisCysIleGlnValAlaGluValLysGluLysThrGlu	1382
Qy	849	CAACTGGCGCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAGCTGAGGGAGGAA	908
Db	1383	LeuLeuGlnThrLeuSerSerAspValSerGluLeuLeuLysAspLysThrHisLeuGln	1402
Qy	909	GAATAACTCAAAGCGGTCACATTACAAATCCAAGGAAGACCGCCAGAGCTGCTCAAGTTA	968
Db	1403	GluLysLeuGlnSerLeu-----GluLysAspSerGlnAlaLeuSerLeuThr	1418
Qy	969	GAAGTGGACTTCGAACACAAGGCCCTCGAGGTTTCCAGGAGCACCGAAGAGATGAACGCC	1028
Db	1419	LysCysGluLeuGluAsnGlnIleAlaGlnLeuAsnLysGluLysGlu-----	1434
Qy	1029	AAATTGGCGAATCAAGAAATCTCACACCGGCAACTTCGACTCAA-----	1073
Db	1435	---LeuLeuValLysGluSerGluSerLeuGlnAlaArgLeuSerGluSerAspTyrGlu	1453
Qy	1074	-----CTGGTT	1079
Db	1454	LysLeuAsnValSerLysAlaLeuGluAlaAlaLeuValGluLysGlyGluPheAlaLeu	1473
Qy	1080	GGCTTATCGCAAAGGATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAG-----	1133
Db	1474	ArgLeuSerSerThrGlnGluGluValHisGlnLeuArgArgGlyIleGluLysLeuArg	1493
Qy	1134	-----GCAGAGGAAGAGCTCCAG---GAGCTGAGAGAGAAAAATTGCCAAGGG	1178
Db	1494	ValArgIleGluAlaAspGluLysLysGlnLeuHisIleAlaGluLysLeuLysGluArg	1513
Qy	1179	GAATGTGGAACCTCCAGTCTCATGGCGGAAGTGAGAGTCTGCGCAAGCGC-----	1229
Db	1514	GluArgGluAsnAspSerLeuLysAspLysValGluAsnLeuGluArgGluLeuGlnMet	1533
Qy	1230	-----GTGCTTGAGATGGAGGGCAAGGATGAAGAGATCACG	1265
Db	1534	SerGluGluAsnGlnGluLeuValIleLeuAspAlaGluAsnSerLysAlaGluValGln	1553
Qy	1266	AAGACCGAGGCCCGAGTCCGGGAGCTGAAGAAGAGCTCCAAGAGGAAGAACCCACAGC	1325
Db	1554	ThrLeuLysThrGlnIleGluGluMetAlaArgSerLeuLysValPheGluLeuAspLeu	1573
Qy	1326	AAGGAACCTTAGACTAGAAGTGGAGAAGCTGCAGAAG-----AGG	1364
Db	1574	ValThrLeuArgSerGluLysGluAsnLeuThrLysGlnIleGlnGluLysGlnGlyGln	1593
Qy	1365	ATGCTGAGCTGGAGAAGCTGGAGGAAGCGTTTCAGCGGAGTAAGTCGGAATGCCACCAG	1424
Db	1594	LeuSerGluLeuAspLysLeuLeuSerSerPheLysSerLeuLeuGluLysGluGln	1613

Qy	1425	CTCCATCTGAACCTCGAGAAAGGAGAAACCTAACCAAAAGACCTGCTG---AACGAGCTG	1481
Db	1614	AlaGluIleGlnIleLysGluGluSerLysThrAlaValGluMetLeuGlnAsnGlnLeu	1633
Qy	1482	GAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAGGCC	1541
Db	1634	LysGluLeuAsnGluAlaValAlaLeu---CysGlyAspGlnGluIleMetLysAla	1652
Qy	1542	-----GAGTTAAGCCTCAAAGATGACCTTACA	1568
Db	1653	ThrGluGlnSerLeuAspProIleGluGluHisGlnLeuArgAsnSerIleGlu	1672
Qy	1569	AAGCTGAAGTCCCTCAGTGTGATGCTGGTGGATGAGAGGAAAAATATG-----ATGGAG	1622
Db	1673	LysLeuArgAla-----ArgLeuGluAlaAspGluLysLysGlnLeuCysValLeuGln	1690
Qy	1623	AAAATAAAGCAAGAAAGAGAGGAAAGTGGATGGGTTCGAATAAAAACTTAAGGTGGAGCAG	1682
Db	1691	GlnLeuLysGluSerGluHisAlaAspLeuLeu-----Lys	1703
Qy	1683	GGAAAAGTCAATGGATGTGACGGGAAAAAGCTA-----	1712
Db	1704	GlyArgValGluAsnLeuGluArgGluLeuGluIleAlaArgThrAsnGlnGluHisAla	1723
Qy	1713	-----ATCGAGGAAAGCAAG-----AAGCTTTTAAAACTCAAATCTGAA---	1751
Db	1724	AlaLeuGluAlaGluAsnSerLysGlyGluValGluThrLeuLysAlaLysIleGluGly	1743
Qy	1752	-----ATGAGGAAAAAGGAGTACAGTCTGACAAAAGGAGAGGGAT	1790
Db	1744	MetThrGlnSerLeuArgGlyLeuGluLeuAspValValThrIleArgSerGluLysGlu	1763
Qy	1791	GAGTGTAGTGGTAAACTGAGGAGCGAAGAAAGAGTCTGTGAACTG-----	1838
Db	1764	AsnLeuThrAsnGluLeuGlnLysGluGlnGluArgIleSerGluLeuGluIleLeAsn	1783
Qy	1838	-----	1838
Db	1784	SerSerPheGluAsnIleLeuGlnGluLysGluGlnGluLysValGlnMetLysGluLys	1803
Qy	1839	---AGCTGCAGTGTAGACTTACTTAAAGAAAGCGGCTTGATGGCATA---GAGGAGGTAGAA	1892
Db	1804	SerSerThrAlaMetGluMetLeuGlnThrGlnLeuLysGluLeuAsnGluArgValAla	1823
Qy	1893	AGGAAATAAACCGAGGTAGTCTGTCGAAG-----GGGTCTGAGTTTC	1934
Db	1824	AlaLeuHisAsnAspGlnGluAlaCysLysAlaLysGluGlnAsnLeuSerSerGlnVal	1843
Qy	1935	ACCTGCCCGGAA-----	1946
Db	1844	GluCysLeuGluLeuGluLysAlaGlnLeuLeuGlnGlyLeuAspGluAlaLysAsnAsn	1863
Qy	1947	-----GACAAT	1952
Db	1864	TyrIleValLeuGlnSerSerValAsnGlyLeuIleGlnGluValGluAspGlyLysGln	1883
Qy	1953	AAGATCAGAGAACTAACGCTTGAAATCGAGAGACTGAAGAAACCGCTCCAGCAGTTGGAG	2012
Db	1884	LysLeuGluLysLysAspGluGluIleSerArgLeuLysAsnGlnIleGlnAspGlnGlu	1903
Qy	2013	GTCGTGGAGGGGACTTGATGAAGACCGAGGACCAATATGAC-----	2054
Db	1904	GlnLeuValSerLysLeuSerGlnValGluGlyGluHisGlnLeuTrpLysGluGlnAsn	1923
Qy	2055	-----CAGTTGGAGCAGAAGTTTCAGAACCGAGCAGGATAAG	2090
Db	1924	LeuGluLeuArgAsnLeuThrValGluLeuGluGlnLysIleGlnValLeuGlnSerLys	1943
Qy	2091	GCAAACTTCCTCTCCAGCAGCTCGAGGAAATCAAAACAC-----	2129
Db	1944	AsnAlaSerLeuGlnAspThrLeuGluValLeuGlnSerSerTyrLysAsnLeuGluAsn	1963
Qy	2130	-----CAAATGCCCAAGCACAAAGCCATAGAGAAA---GGGAGGCCCGTG	2171

Db 1964 GluLeuGluLeuThrLysMetAspLysMetSerPheValGluLysValAsnLysMetThr 1983
QY 2172 AGCCAGGAAGCCGAACTCGCACACAGAGTTTCGGCTGGAGGAGCTAAAGTCGTGATTTA 2231
Db 1984 AlaLysGluThrGluLeuGlnArgGluMetHisGluMetAlaGlnLysThrAlaGluLeu 2003
QY 2232 CAGGCCGAGGTGCGAGGCTCTCAAGGAGAAG-----ATCCAC 2267
Db 2004 GlnGluGluLeuSerGlyGluLysAsnArgLeuAlaGlyGluLeuGlnLeuLeuGlu 2023
QY 2268 GAGCTGATGAACAAGAACAGACAGCTGTCTCAGCTCCAAGTCGACTATTTCGGTCTTCAG 2327
Db 2024 GluIleLysSerSerLysAspGlnLeuLysGluLeuThrLeuGluAsnSerGluLeuLys 2043
QY 2328 CAAAGATT-----ATGGAAGAAGAACTAAGAACAAAGAACATGGGGAGGGAG 2375
Db 2044 LysSerLeuAspCysMetHisLysAspGlnValGluLysGluGlyLysValArgGluGlu 2063
QY 2376 GTCCTCAATCTGACCAAGGAGCTAGAGCTTTCGAAGCGCTACAGCCGAGCTCTCAGGCCG 2435
Db 2064 IleAlaGluTyrGlnLeuArgLeuHisGluAlaGluLysLysHisGlnAlaLeuLeu 2083
QY 2436 AGTGGGAACGGCCGAAGGATGTTGGACGTGCCTGCGCTCCACTGGGGTCAGACCGAG 2495
Db 2084 AspThrAsnLysGlnTyrGluValGluIleGlnThrTyrArgGluLysLeuThrSerLys 2103
QY 2496 GCGGTGTGGGGATGCTGCGGAGGAGGAGACCCCGCTGTGTTTCATTCGAAATCCTTC 2555
Db 2104 GluGluCysLeuSerSerGlnLysLeuGlu-----IleAspLeuLeuLysSerSer 2120
QY 2556 CAGGAGGAA-----AATCACATC----- 2573
Db 2121 LysGluGluLeuAsnAsnSerLeuLysAlaThrThrGlnIleLeuGluGluLeuLysLys 2140
QY 2574 -----ATGAGTAATCTTCGACAGGTAGGC---CTGAAGAAACCCATGGAACGGTCC--- 2621
Db 2141 ThrLysMetAspAsnLeuLysTyrValAsnGlnLeuLysLysGluAsnGluArgAlaGln 2160
QY 2622 TCGGTCTCGACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAATCTTGGATT 2681
Db 2161 GlyLysMetLysLeuLeuIleLysSerCysLysGlnLeuGluGluLysGlu----- 2178
QY 2682 CCTTGGATGAGAAAAGAGAAAACGGTCCCTCCACTCCGAGGAGAAAGGCCAGGCCA 2741
Db 2179 ---IleLeuGlnLysGluLeuSerGlnLeuGlnAlaAlaGlnGluLys----- 2193
QY 2742 AACAGGGTGCAGGGCACCCCGGGAGCTGCTCTAGCACCAAAGCAGGGCCAGCCCTA 2801
Db 2194 -----GlnLysThrGlyThrValMet 2200
QY 2802 CACATCCGTGTG-----ACACAGATCATGAGAACAGCACTGCCACCTGGAGATC 2852
Db 2201 AspThrLysValAspGluLeuThrThrGluIleLysGluLeuLysGluThrLeuGluGlu 2220
QY 2853 ACAAGCCCCACATCTGAAGAGTTTCTCTAGT----- 2885
Db 2221 LysThrLysGluAlaAspGluTyrLeuAspLysTyrCysSerLeuLeuIleSerHisGlu 2240
QY 2886 -----ACCACCGTCATTCCTACCTTACCTTAGCAACACAGAAA 2918
Db 2241 LysLeuGluLysAlaLysGluMetLeuGluThrGlnValAlaHisLeuCysSerGlnGln 2260
QY 2919 CCAAGAATAACCATATTATTCATCACCAATGTGTCGAAAAGCCCAAAAGTGCAGAT 2978
Db 2261 -----SerLysGlnAspSerArgGlySer 2268
QY 2979 CCTACTCTCGGCCAGAACAGCCATGTCCCTGTCTCAGCATTAATACTATTTCCAGAGAG 3038
Db 2269 ProLeuLeuGlyProVal---ValProGlyProSerProIleProSerValThrGluLys 2287
QY 3039 AAGAGCCCGGAGGTGAAGGAGCGCTTTGCCGACAGG----- 3077

Db 2288 ArgLeuSerSerGlyGlnAsnLysAlaSerGlyLysArgGlnArgSerSerGlyIleTrp 2307
QY 3078 -----CCTGCATCCCCCATCCAAATCATGACGGTGTCAACATCT 3116
Db 2308 GluAsnGlyGlyGlyProThrProAlaThrProGluSerPheSerLysLysSerLysLys 2327
QY 3117 GCAGCTCCCAATCGCTGTCTCTCTCGAATCTCAG----- 3155
Db 2328 AlaValMetSerGlyIleHisProAlaGluAspThrGluGlyThrGluPheGluProGlu 2347
QY 3156 -----GAAGTGCCTATGGGAAGGACT----- 3176
Db 2348 GlyLeuProGluValValLysLysGlyPheAlaAspIleProThrGlyLysThrSerPro 2367
QY 3177 -----ATCCTCAAAGTCACCCCGGAAAAACAACTGTTCCAGCCCCCGTGGGAAGTAC 3230
Db 2368 TyrIleLeuArgArgThrThrMetAlaThrArgThrSerProArgLeuAlaGlnLys 2387
QY 3231 AACTCCAATGCTAATATCATCACCCGGAAGACAATAAAATTCACATTACCTGGGTTCT 3290
Db 2388 LeuAlaLeuSerProLeuSerLeuGlyLysGluAsn----- 2399
QY 3291 CAGTTTAAGCGATCTCTGGGCTGCCGCTGAAAGCGTGAGCCCGAGTTATCACCGTCCGG 3350
Db 2400 ---LeuAlaGluSerSerLysProThrAlaGlyGlySerArgSer-----Gln 2414
QY 3351 CCTGTCAACGTGACAGCGGAGAGAGGAGTTTCTACAGGCACAGTCCTTCGCTCTCCAGG 3410
Db 2415 LysValLysValAlaGlnArgSerProValAspSerGlyThrIleLeuArgGluProThr 2434
QY 3411 -----AACCACCTC-----TCITCAAGACCC 3431
Db 2435 ThrLysSerValProValAsnAsnLeuProGluArgSerProThrAspSerProArgGlu 2454
QY 3432 GGTGTAGCAAGTGACCACTATACTATAACCCCGGTCAACCGTCATCCACACGA 3491
Db 2455 GlyLeuArgValLysArgGlyArgLeuValProSerProLysAlaGlyLeuGluSerLys 2474
QY 3492 GGAACCAATCAGTGTGAGGACAA 3515
Db 2475 GlySerGluAsnCysLysValGln 2482

RESULT 5
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:

```
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

Alignment Scores:
Pred. No.: 1,17e-21 Length: 3248
Score: 436.50 Matches: 316
Percent Similarity: 36.27% Conservative: 240
Best Local Similarity: 20.61% Mismatches: 527
Query Match: 5.65% Indels: 450
DB: 1 Gaps: 63

US-10-788-793-1 (1-4364) x US-08-353-700-1 (1-3248)

QY 51 TTAAGGAGTCGACAAACAGGTGGGAATGAGATCAGCAAAAT-----CAAGGTGGAGAAAGT 104
Db 1746 LeuSerSerArgSerLeuLeuGlyIleAspThrGluAspAlaIleGlnGlyArgAsnGlu 1765
QY 105 TCA-----TCTAACGGGCATGTCCTCTGCCCCCAAGTCCTCCATCATCAGCAGTGAT 155
Db 1766 SerCysAspIleSerLysGluHisThrSerGluThrThrGluArgThrProLysHisAsp 1785
QY 156 GGTGTAAGGGCCCCCTCAGAAGATGCAAAAAAGAACAGGCCAATCGGAAGGAGGAGAT 215
Db 1786 ValHisGlnIleCysAspLysAspAlaGlnGlnAsp---LeuAsnLeuAspIleGluLys 1804
QY 216 GTCATGGCTTCGGAACTATCAAAAGGCACCTCAAACCATCTGGAGAA-----AGTGAG 269
Db 1805 IleThrGluThrGlyAlaVal-----LysProThrGlyGluCysSerGlyGlu 1820
QY 270 AAAAGACTAAGAAGTCTGTGGAGTTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATC 329
Db 1821 GlnSerProAspThrAsnTyrGluProGlyGluAspLysThrGlnGlySerSerGlu 1840
QY 330 ATGGAAGGGAGTTGCAG-----GCT 350
Db 1841 CysIleSerGluLeuSerPheSerGlyProAsnAlaLeuValProMetAspPheLeuGly 1860
QY 351 CGAGAAAGATGTCATCCAC-----ATGCTGAGGACAGAGAAAAACCAAGCCCGAGGTCTGTG 404
Db 1861 AsnGlnGluAspIleHisAsnLeuGlnLeuArgValLysGluThrSerAsnGluAsnLeu 1880
QY 405 GAG---GCACACTATGGATCTGCAGAACCTGAGAAAGTGCTTCGGGTCTGCACCGA--- 458
Db 1881 ArgLeuLeuHisValIleGluAspArgAspArgLysValGluSerLeuLeuAsnGluMet 1900
QY 459 -----GATGCC 464
Db 1901 LysGluLeuAspSerLysLeuHisLeuGlnGluValGlnLeuMetThrLysIleGluAla 1920
QY 465 ATCCTTGCTCAAGAGAGTCCATAGGAGAAAGACGTCTATGAGAAACCTATCTCAGAGCTG 524
Db 1921 CysIleGluLeuGluLysIleValGlyGlu-----LeuLysLysGluAsnSerAspLeu 1938
QY 525 GACAGACTGGAGGAAAAGCAGAGGACAGCTACCGCCGCATGCTAGAGCAGCTGCTGCTG 584
Db 1939 SerGluLysLeuGluTyrPheSerCysAspHisGlnGluLeuLeuGlnArgValGluThr 1958
QY 585 GCTGAGAAAGTGTACAGGCGCACCGTGTACGAGCTGGAGAACGAGAAAGCACAAAGCACACT 644
Db 1959 SerGlu-----GlyLeuAsnSerAspLeuGluMetHisAla 1970
```

```
QY 645 GACTACATGAACAAGACGACGACTTCACCAACCTG----- 680
Db 1971 AspLysSerSerArgGluAspIleGlyAspAsnValAlaLysValAsnAspSerTirpLys 1990
QY 681 -----CTGGAGCAGGAGCGCAGAGAGGTTG-----AAAAAGCTCCTT 716
Db 1991 GluArgPheLeuAspValGluAsnGluLeuSerArgIleArgSerGluLysAlaSerIle 2010
QY 717 GAACAAGAAAAAGCTTACCAA----- 737
Db 2011 GluHisGluAlaLeuTyrLeuGluAlaAspLeuGluValValGlnThrGluLysLeuCys 2030
QY 738 GCCCGCAAGAAAAAGGAAACGCTAAGCGGCTC----- 770
Db 2031 LeuGluLysAspAsnGluAsnLysGlnLysValIleValCysLeuGluGluLeuSer 2050
QY 771 -----AACAAACTTCGAGATGAGCTT-----GTGAAGCTCAAGTCC 806
Db 2051 ValValThrSerGluArgAsnGlnLeuArgGlyGluLeuAspThrMetSerLysLysThr 2070
QY 807 TTCGCCCTCATGTTGGTGGACGAGAGGCAG----- 836
Db 2071 ThrAlaLeuAspGlnLeuSerGluLysMetLysGluLysThrGlnGluLeuGluSerHis 2090
QY 837 -----ATGCAC-----ATCGAG 848
Db 2091 GlnSerGluCysLeuHisCysIleGlnValAlaGluAlaGluValLysGluLysThrGlu 2110
QY 849 CAACTGGGCCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAAGCTCAGAGGGAGGAA 908
Db 2111 LeuLeuGlnThrLeuSerSerAspValSerGluLeuLeuLysAspLysThrHisLeuGln 2130
QY 909 GAAAAACTCAAAGCGGTCACCTTACAAATCCAAAGGAGACCGCCAGAAAGCTGTCTCAAGTTA 968
Db 2131 GluLysLeuGlnSerLeu-----GluLysAspSerGlnAlaLeuSerLeuThr 2146
QY 969 GAAGTGGACTTCGAACACAAAGGCCTCGAGGTTTCCAGGAGCAGCAGAGAGATGAACGCC 1028
Db 2147 LysCysGluLeuGluAsnGlnIleAlaGlnLeuAsnLysGluLysGlu----- 2162
QY 1029 AATTGGCGAATCAAGAATCTCAACAACCGGCAACTTCGACTCAA----- 1073
Db 2163 --LeuLeuValLysGluSerGluSerLeuGlnAlaArgLeuSerGluSerAspTyrGlu 2181
QY 1074 -----CTGGTT 1079
Db 2182 LysLeuAsnValSerLysAlaLeuGluAlaLeuValGluLysGlyGluPheAlaLeu 2201
QY 1080 GGCTTATCGCAACGATTGAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAG----- 1133
Db 2202 ArgLeuSerSerThrGlnGluGluValHisGlnLeuArgArgGlyIleGluLysLeuArg 2221
QY 1134 -----GCAGAGGAAGAGCTCCAG---GAGCTGAGAGAGAAAAATTGCCAAAGGG 1178
Db 2222 ValArgIleGluAlaAspGluLysLysGlnLeuHisIleAlaGluLysLeuLysGluArg 2241
QY 1179 GAATGTGGAAACTCCAGTCTCATGCGGGAAGTGGAAGTCTGCCGAAGCGC----- 1229
Db 2242 GluArgGluAsnAspSerLeuLysAspLysValGluAsnLeuGluArgGluLeuGlnMet 2261
QY 1230 -----GTGCTTGAGATGGAGGGCAAGGATGAAGAGATCAGC 1265
Db 2262 SerGluGluAsnGlnGluLeuValIleLeuAspAlaGluAsnSerLysAlaGluValGlu 2281
QY 1266 AAGACCGAGGCCAGTGC CGGAGCTGAAGAAGAGCTCCAAGAGGAAGAACACACAGC 1325
Db 2282 ThrLeuLysThrGlnIleGluGluMetAlaArgSerLeuLysIlePheGluLeuAspLeu 2301
QY 1326 AAGAACTTAGACTAGAAGTGGAGAGCTGCAGAAAG-----AGG 1364
Db 2302 ValThrLeuArgSerGluLysGluAsnLeuThrLysGlnIleGlnGluLysGlnGlyGln 2321
QY 1365 ATGCTGAGCTGAGAGAGCTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGCACCCAG 1424
```

Db	2322	LeuSerGluLeuAspLysLeuLeuSerPheLysSerLeuLeuGluGluLysGluGln	2344
QY	1425	CTCCATCTGAACCTGGAGACGAGAGAACCTAACCAAAAGACCTGCTG---AACGAGCTG	1481
Db	2342	AlaGluIleGlnIleLysGluSerLysThrAlaValGluMetLeuGlnAsnGlnLeu	2361
QY	1482	GAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAGGCC	1541
Db	2362	LysGluLeuAsnGluAlaValAlaAlaLeu---CysGlyAspGlnGluIleMetLysAla	2380
QY	1542	-----GAGTTAAGCCTCAAAGATGACCTTACA	1568
Db	2381	ThrGluGlnSerLeuAspProIleGluGluGlnHisGlnLeuArgAsnSerIleGlu	2400
QY	1569	AAGCTGAAGTCCTTCACCTGCTGATGCTGGTGATGAGAGGAAAAATATG-----ATGGAG	1622
Db	2401	LysLeuArgAla-----ArgLeuGluAlaAspGluLysLysGlnLeuCysValLeuGln	2418
QY	1623	AAATAAAGCAAGAGAGAGGAAAGTGGATGGGTGTAATAAAAACTTTAAGGTGGAGCAG	1682
Db	2419	GlnLeuLysGluSerGluHisAlaAspLeuLeu-----Lys	2431
QY	1683	GGAAAAATCATGGATGTGACGAAAAAGCTA-----	1712
Db	2432	GlyArgValGluAsnLeuGluArgGluLeuGluIleAlaArgThrAsnGlnGluHisAla	2451
QY	1713	-----ATCGAGGAACCAAG-----AAGCTTTTAAAACTCAAATCTGAA---	1751
Db	2452	AlaLeuGluAlaGluAsnSerLysGlyGluValGluThrLeuLysAlaLysIleGluGly	2471
QY	1752	-----ATGGAGGAAAAAGGAGTACAGTCTGACAAAAGGAGAGGGAT	1790
Db	2472	MetThrGlnSerLeuArgGlyLeuGluLeuAspValValThrIleArgSerGluLysGlu	2491
QY	1791	GAGCTGATGGTAAACTGAGGAGCGAAGAAGGTCCTGTGAACTG-----	1838
Db	2492	AsnLeuThrAsnGluLeuGlnLysGluGlnGluArgIleSerGluLeuGluIleLeuAsn	2511
QY	1838	-----	1838
Db	2512	SerSerPheGluAsnIleLeuGlnGluLysGluGlnGluLysValGlnMetLysGluLys	2531
QY	1839	---AGCTGCAGTGTAGACTTACTAAAGAAGCGGCTTGATGGCATA--GAGGAGGTAGAA	1892
Db	2532	SerSerThrAlaMetGluMetLeuGlnThrGlnLeuLysGluLeuAsnGluArgValAla	2551
QY	1893	AGGGAATAAACCAGGTAGGTGTCGCAAG-----GGGTCTGAGTTC	1934
Db	2552	AlaLeuHisAsnAspGlnGluAlaCysLysAlaLysGluGlnAsnLeuSerSerGlnVal	2571
QY	1935	ACCTGCCCGGAA-----	1946
Db	2572	GluCysLeuGluLeuGluLysAlaGlnLeuLeuGlnGlyLeuAspGluAlaLysAsnAsn	2591
QY	1947	-----GACAAT	1952
Db	2592	TyrIleValLeuGlnSerSerValLysGlyLeuIleGlnGluValGluAspGlyLysGln	2611
QY	1953	AAGATCAGAGAACTAACGCTTGAAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAG	2012
Db	2612	LysLeuGluLysLysAspGluGluIleSerArgLeuLysAsnGlnIleGlnAspGlnGlu	2631
QY	2013	GTGGTGGAGGGGACTTGTGATGAAGACCGGACCGCAATATGAC-----	2054
Db	2632	GlnLeuValSerLysLeuSerGlnValGluGlyGluHisGlnLeuTrpLysGluGlnAsn	2651
QY	2055	-----CAGTTGGAGCAGAAGTTCCAGAACCGGACGAGGATAAG	2090
Db	2652	LeuGluLeuArgAsnLeuThrValGluLeuGluGlnLysIleGlnValLeuGlnSerLys	2671
QY	2091	GCAAACTTCCTCTCCACGACGCTCGAGGAAATCAAAAC-----	2129

2672	AsnAlaSerLeuGlnAspThrLeuGluValLeuGlnSerSerTyrLysAsnLeuGluAsn	2691
2130	-----CAAATGGCCCAAGCACAAAGCCATAGAGAAA--GGGAGGCCGTG	2171
2692	GluLeuGluLeuThrLysMetAspLysMetSerPheValGluLysValAsnLysMetThr	2711
2172	AGCCAGGAAGCCGAACACTGCGACACACAGGTTTCGGCTGGAGGAGGCTAAAGTTCGTGATT	2231
2712	AlaLysGluThrGluLeuGlnArgGluMetHisGluMetAlaGlnLysThrAlaGluLeu	2731
2232	CAGGCCGAGGTGCAGCTCTCAAGGAGAAG-----ATCCAC	2267
2732	GlnGluGluLeuSerGlyGluLysAsnArgLeuAlaGlyGluLeuGlnLeuLeuGlu	2751
2268	GAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTATTCGGTCTCTCAG	2327
2752	GluLeuLysSerSerLysAspGlnLeuLysGluLeuThrLeuGluAsnSerGluLeuLys	2771
2328	CAAAGATT-----ATGGAAGAGAAACTAAGAACAAAGAACATCGGGAGGGAG	2375
2772	LysSerLeuAspCysMetHisLysAspGlnValGluLysGluGlyLysValArgGluGlu	2791
2376	GTCTCTCAATCTGACCAAGGAGCTAGAGCTTTCGAAGCGCTACAGCCGAGCTCTCAGGCCG	2435
2792	IleAlaGluTyrGlnLeuArgLeuHisGluAlaGluLysLysHisGlnAlaLeuLeu	2811
2436	AGTGGGAACGGCCGAAGGATGGTGGACGTGCCTGTGGCCTCCACTGGGGTGCAGACCGAG	2495
2812	AspThrAsnLysGlnTyrGluValGluIleGlnThrTyrArgGluLysLeuThrSerLys	2831
2496	GCGGTGTGCGGGGATCTGCGGAGGAGGAGACCCCGCTGTGTTCATTCGCAAAATCCTTC	2555
2832	GluGluCysLeuSerSerGlnLysLeuGlu-----IleAspLeuLeuLysSerSer	2848
2556	CAGGAGGAA-----AATCACATC-----	2573
2849	LysGluGluLeuAsnAsnSerLeuLysAlaThrThrGlnIleLeuGluGluLysLys	2868
2574	-----ATGAGTAATCTTCGACAGGTAGGC--CTGAAGAAACCCATGGAACGGTCC--	2621
2869	ThrLysMetAspAsnLeuLysTyrValAsnGlnLeuLysLysGluAsnGluArgAlaGln	2888
2622	TCGGTCTCGACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCTTGGATT	2681
2889	GlyLysMetLysLeuLeuIleLysSerCysLysGlnLeuGluGluLysGlu-----	2906
2682	CCTTGGATGAGAAAAAGAGAAAACCGGTCTCTCCACTCCGCGAGGAGAAAGGCCAGGCCA	2741
2907	---IleLeuGlnLysGluLeuSerGlnLeuGlnAlaAlaGlnGluLys-----	2921
2742	AACCAGGTGCAGGGCACCCCGGGGAGCTGGTCTTAGCACCAAGCAGGGGCCAGCCCTTA	2801
2922	-----GlnLysThrGlyThrValMet	2928
2802	CACATCCGTGTG-----ACACCAGATCATGAGAACAGCACTGCCACCCCTGGAGATC	2852
2929	AspThrLysValAspGluLeuThrThrGluIleLysGluLeuLysGluThrLeuGluGlu	2948
2853	ACAAGCCCCACATCTGAAGAGT-----TAGT-----	2885
2949	LysThrLysGluAlaAspGluTyrLeuAspLysTyrCysSerLeuLeuIleSerHisGlu	2968
2886	-----ACCACCGTCATTCCTACCTTAGGCAACCCAGAGAAA	2918
2969	LysLeuGluLysAlaLysGluMetLeuGluThrGlnValAlaHisLeuCysSerGlnGln	2988
2919	CCAAGAATAACCATTTATCCATCACCCCAATGTATGTGCGCAAAAGCCCAAAAGTCAGAT	2978
2989	-----SerLysGlnAspSerArgGlySer	2996
2979	CCTACTCTCGGCCCA--GAACGAGCCATGTCCCTGTACAGTACTACTACTATTTCCAGA	3035
2997	ProLeuLeuGlyProValValProGlyProSerProIleProSerValThrGluLysArg	3016

Db 1971 AspLysSerSerArgGluAspIleGlyAspAsnValAlaLysValAsnAspSerTrpLys 1990
QY 681 -----CTGGAGCAGGAGCGAGAGCTTG-----AAAAAGCTCCTT 716
Db 1991 GluArgPheLeuAspValGluAsnGluLeuSerArgIleArgSerGluLysAlaSerIle 2010
QY 717 GAACAAGAAAAAGCTTACCAA----- 737
Db 2011 GluHisGluAlaLeuTyrLeuGluAlaAspLeuGluValValGlnThrGluLysLeuCys 2030
QY 738 GCCCGCAAGAAAGGAAAAACGCTAAGCGGCTC----- 770
Db 2031 LeuGluLysAspAsnGlnAsnLysGlnLysValIleValCysLeuGluGluLeuSer 2050
QY 771 -----AACAAACTTCGAGATGAGCTT-----GTGAAGCTCAAGTCC 806
Db 2051 ValValThrSerGluArgAsnGlnLeuArgGlyGluLeuAspThrMetSerLysLysThr 2070
QY 807 TTCGCCCTCATGTTGGTCGACGAGGCGAG----- 836
Db 2071 ThrAlaLeuAspGlnLeuSerGluLysMetLysGluLysThrGlnGluLeuGluSerHis 2090
QY 837 -----ATGCAC-----ATCGAG 848
Db 2091 GlnSerGluCysLeuHisCysIleGlnValAlaGluValLysGluLysThrGlu 2110
QY 849 CAACTGGCCCTGCAGAGTCAGAAAGTCCAGGACCTCACTCAGAAAGCTGAGGAGGAGAA 908
Db 2111 LeuLeuGlnThrLeuSerSerAspValSerGluLeuLeuLysAspLysThrHisLeuGln 2130
QY 909 GAAAAACTCAAAGCGGTCACCTTACAAATCCAAAGGAAGACCGCCAGAAAGCTGCTCAAGTTA 968
Db 2131 GluLysLeuGlnSerLeu-----GluLysAspSerGlnAlaLeuSerLeuThr 2146
QY 969 GAAGTGGACTTCGAACACAAAGGCCTCGAGGTTTCCAGGAGCACCAGAGATGAACGCC 1028
Db 2147 LysCysGluLeuGluAsnGlnIleAlaGlnLeuAsnLysGluLysGlu----- 2162
QY 1029 AAATTGGCGAATCAAGAACTCTCAACACCGGCAACTTCGACTCAAA----- 1073
Db 2163 ---LeuLeuValLysGluSerGluSerLeuGlnAlaArgLeuSerGluSerAspTyrGlu 2181
QY 1074 -----CTGGTT 1079
Db 2182 LysLeuAsnValSerLysAlaLeuGluAlaAlaLeuValGluLysGlyGluPheAlaLeu 2201
QY 1080 GGCTTATCGCAAGGATTGAGGAGCTGGAAGAGACCAATAAAGCCTTCAGAAAG----- 1133
Db 2202 ArgLeuSerSerThrGlnGluGluValHisGlnLeuArgArgGlyIleGluLysLeuArg 2221
QY 1134 -----GCAGAGAAAGACTCCAG-----GAGCTGAGAGAGAAATAATGCCAAAGGG 1178
Db 2222 ValArgIleGluAlaAspGluLysLysGlnLeuHisIleAlaGluLysLeuLysGluArg 2241
QY 1179 GAATGTGGAACCTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGC----- 1229
Db 2242 GluArgGluAsnAspSerLeuLysAspLysValGluAsnLeuGluArgGluLeuGlnMet 2261
QY 1230 -----GTGCTTGAGATGGAGGCGCAAGGATGAAGAGATCAGC 1265
Db 2262 SerGluGluAsnGlnGluLeuValIleLeuAspAlaGluAsnSerLysAlaGluValGlu 2281
QY 1266 AAGACCGAGGCCAGTGCCGGGAGCTGAAGAAGAGCTCCAAGAGGAAGAACACACAGC 1325
Db 2282 ThrLeuLysThrGlnIleGluGluMetAlaArgSerLeuLysIlePheGluLeuAspLeu 2301
QY 1326 AAGGAACCTTAGACTAGAGTGGAGAGCTGCAGAAG-----AGG 1364
Db 2302 ValThrLeuArgSerGluLysGluAsnLeuThrLysGlnIleGlnGluLysGlnGlyGln 2321
QY 1365 ATGCTGTAGCTGGAGAGCTGGAGGAAGCGTTTACGCCGAGTAAGTCGGAATGCACCCAG 1424
Db 2322 LeuSerGluLeuAspLysLeuLeuSerSerPheLysSerLeuLeuGluGluLysGluGln 2341

QY 1425 CTCCATCTGAACCTGGAGAAGGAGAAGAACTAAACCAAAGACCTGCTG----AACGAGCTG 1481
Db 2342 AlaGluIleGlnIleLysGluGluSerLysThrAlaValGluMetLeuGlnAsnGlnLeu 2361
QY 1482 GAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAGGCC 1541
Db 2362 LysGluLeuAsnGluAlaValAlaAlaLeu---CysGlyAspGlnGluIleMetLysAla 2380
QY 1542 -----GAGTTAAGCCTCAAAGATGACCTTACA 1568
Db 2381 ThrGluGlnSerLeuAspProIleGluGluGluHisGlnLeuArgAsnSerIleGlu 2400
QY 1569 AAGCTGAAGTCCTTCACTGTGTGCTGCTGGATGAGAGGAAAAATATG-----ATGGAG 1622
Db 2401 LysLeuArgAla-----ArgLeuGluAlaAspGluLysGlnLeuCysValLeuGln 2418
QY 1623 AAAATAAAGCAAGAGAGAGAGAAAGTGGATGGGTTGAATAAAAACTTTAAGGTGGAGCAG 1682
Db 2419 GlnLeuLysGluSerGluHisHisAlaAspLeuLeu-----Lys 2431
QY 1683 GGAAGAGTCATGGATGTGACGGAAGGCTA----- 1712
Db 2432 GlyArgValGluAsnLeuGluArgGluLeuGluIleAlaArgThrAsnGlnGluHisAla 2451
QY 1713 -----ATCGAGGAAAGCAAG-----AAGCTTTTAAAACTCAAATCTGAA--- 1751
Db 2452 AlaLeuGluAlaGluAsnSerLysGlyGluValGluThrLeuLysAlaLysIleGluGly 2471
QY 1752 -----ATGGAGGAAAAAGGAGTACAGTCTGACAAAGAGAGGGAT 1790
Db 2472 MetThrGlnSerLeuArgGlyLeuGluLeuAspValValThrIleArgSerGluLysGlu 2491
QY 1791 GAGCTGATGGGTAAACTGAGGAGCGGAAGAAAGGTCTCTGTGAACCTG----- 1838
Db 2492 AsnLeuThrAsnGluLeuGlnLysGluGlnGluArgIleSerGluLeuGluIleIleAsn 2511
QY 1838 ----- 1838
Db 2512 SerSerPheGluAsnIleLeuGlnGluLysGluGlnGluLysValGlnMetLysGluLys 2531
QY 1839 ---AGCTGCAGTGTAGACTTACTAAAGAAAGCGGCTTGATGGCATA---GAGGAGGTAGAA 1892
Db 2532 SerSerThrAlaMetGluMetLeuGlnThrGlnLeuLysGluLeuAsnGluArgValAla 2551
QY 1893 AGGGAATAAACCGAGGTAGTCTGTCGAAG-----GGTCTGAGTTT 1934
Db 2552 AlaLeuHisAsnAspGlnGluAlaCysLysAlaLysGluGlnAsnLeuSerSerGlnVal 2571
QY 1935 ACCTGCCCGGAA----- 1946
Db 2572 GluCysLeuGluLeuGluLysAlaGlnLeuLeuGlnGlyLeuAspGluAlaLysAsnAsn 2591
QY 1947 -----GACAAT 1952
Db 2592 TyrIleValLeuGlnSerSerValLysGlyLeuIleGlnGluValGluAspGlyLysGln 2611
QY 1953 AAGATCAGAGAACTAACGCTTGAAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAG 2012
Db 2612 LysLeuGluLysLysAspGluGluIleSerArgLeuLysAsnGlnIleGlnAspGlnGlu 2631
QY 2013 GTGGTGGAGGGGACTTGATGAAGACCGACCGACCAATATGAC----- 2054
Db 2632 GlnLeuValSerLysLeuSerGlnValGluGlyGluHisGlnLeuTrpLysGluGlnAsn 2651
QY 2055 -----CAGTTGGAGCAGAAAGTTTCAGAACCCAGCAGGATAAG 2090
Db 2652 LeuGluLeuArgAsnLeuThrValGluLeuGluGlnLysIleGlnValLeuGlnSerLys 2671
QY 2091 GCAAACTTCTCTCCAGCAGCTCGAGGNAATCAAAACAC----- 2129
Db 2672 AsnAlaSerLeuGlnAspThrLeuGluValLeuGlnSerSerTyrLysAsnLeuGluAsn 2691

QY	2130	-----CAATGGCCAAAGCACAAGCCATAGAGAAA--GGGAGGCGCGTG	2171
Db	2692	GluleuGluLeuThrLysMetAspLysMetSerPheValGluLysValAsnLysMetThr	2711
QY	2172	AGCCAGGAAGCCGAACCTCGACACACAGGTTTCGGCTGGAGGAGCTAAAGTCGTGATT	2231
Db	2712	AlaLysGluThrGluLeuGlnArgGluMetHisGluMetAlaGlnLysThrAlaGluLeu	2731
QY	2232	CAGGCGGAGGTGCAGGCTCTCAAGGAGAAG-----ATCCAC	2267
Db	2732	GlnGluLeuSerGlyLysAsnArgLeuAlaGlyGluLeuGlnLeuLeuGlu	2751
QY	2268	GAGCTGATGAACAAGGAAGACAGCTGTCTCAGCTCCAAGTCGACTATTCGGTCCCTCAG	2327
Db	2752	GluleLysSerSerLysAspGlnLeuLysGluLeuThrLeuGluAsnSerGluLeuLys	2771
QY	2328	CAAGATT-----ATGGAAGAAGAACTAAGAACCAAGACATGGGAGGGAG	2375
Db	2772	LysSerLeuAspCysMetHisLysAspGlnValGluLysGluGlyLysValArgGluGlu	2791
QY	2376	GTCCTCAATCTGACCAAGGAGCTAGAGCTTTCCAAGCGCTACAGCGAGCTCTCAGGCGG	2435
Db	2792	IleAlaGluTyrGlnLeuArgLeuHisGluAlaGluLysLysHisGlnAlaLeuLeu	2811
QY	2436	AGTGGAAACGGCCGAAGATGGTGGACGTGTGTGGCTTCCACTGGGGTGCAGACCGAG	2495
Db	2812	AspThrAsnLysGlnTyrGluValGluIleGlnThrTyrArgGluLysLeuThrSerLys	2831
QY	2496	CGCGTGTGGGGGATGCTCGGAGGAGGAGACCCCGGCTGTGTTCATTCGCAATCTTC	2555
Db	2832	GluGluCysLeuSerSerGlnLysLeuGlu-----IleAspLeuLeuLysSerSer	2848
QY	2556	CAGGAGGAA-----AATCACATC-----	2573
Db	2849	LysGluGluLeuAsnAsnSerLeuLysAlaThrThrGlnIleLeuGluLeuLysLys	2868
QY	2574	-----ATGAGTAATCTTCGACAGGTAGGC---CTGAAGAAACCCATGGAACGGTCC---	2621
Db	2869	ThrLysMetAspAsnLeuLysTyrValAsnGlnLeuLysLysGluAsnGluArgAlaGln	2888
QY	2622	TCGGTCTCGACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCCTGGATT	2681
Db	2889	GlyLysMetLysLeuLeuLysSerCysLysGlnLeuGluGluLysGlu-----	2906
QY	2682	CCITGGATGAGAAAAGAGAAAACGGTCTTCCACTCCGCAGGAGAAAGGGCCAGGCCA	2741
Db	2907	---IleLeuGlnLysGluLeuSerGlnLeuGlnAlaAlaGlnGluLys-----	2921
QY	2742	AACCAGGTGCAGGCAACCCCGGAGCTGGTCTTAGCACCAAGCAGGGCCAGCCCTTA	2801
Db	2922	-----GlnLysThrGlyThrValMet	2928
QY	2802	CACATCCGTGTG-----ACACCAGATCATGAGAAACAGCACTGCCACCTCGAGATC	2852
Db	2929	AspThrLysValAspGluLeuThrThrGluIleLysGluLeuLysGluThrLeuGluGlu	2948
QY	2853	ACAAGCCCCACATCTGAAGAGTTTCTCTAGT-----	2885
Db	2949	LysThrLysGluAlaAspGluTyrLeuAspLysTyrCysSerLeuLeuIleSerHisGlu	2968
QY	2886	-----ACCACCGTCAATCTTACCTTAGGCAACACAGAAA	2918
Db	2969	LysLeuGluLysAlaLysGluMetLeuGluThrGlnValAlaHisLeuCysSerGlnGln	2988
QY	2919	CCAAGAATAACCATTAATCCATCACCCCAATGTTCATGTGCGAAAAGCCCCAAGTGCAGAT	2978
Db	2989	-----SerLysGlnAspSerArgGlySer	2996
QY	2979	CCTACTCTCGGCCCCA---GAACGAGCCATGTCCCTGTCTCAGCATTAATTTCCAGA	3035
Db	2997	ProLeuLeuGlyProValValProGlyProSerProIleProSerValThrGluLysArg	3016
QY	3036	GAGAAAGCCCGAAGGTGGAAGGAGCGCCTTTCGCGACAGGCTGCATCCCCATCCAA	3095

[illegible]

```
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-056-200-94

Alignment Scores:
Pred. No.: 3.17e-21 Length: 1898
Score: 429.00 Matches: 216
Percent Similarity: 41.20% Conservative: 168
Best Local Similarity: 23.18% Mismatches: 376
Query Match: 5.55% Indels: 172
DB: 1 Gaps: 35

US-10-788-793-1 (1-4364) x US-08-056-200-94 (1-1898)
QY 174 GAAGATGCAAAAAGAACAAAGGCCAATCGGAAGGAGGAGGATGTCATGGCTTCGGAACT 233
Db 344 GluGluArgArgGluGlnGlnLeuArgArgGluGlnGlnGluGluArgArgGluGlnGln 363
QY 234 ATCAAAAGGCACCTCAAAACCATCTGGAGAAAGTGAGAAAAAGACTAAGAAGTCTGTGGAG 293
Db 364 LeuArgArgGluGlnGlnGluGluArgArgGluGlnGlnLeuArgArgGluGlnGluGlu 383
QY 294 TTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGAGTTGAGGCTCGA 353
Db 384 GluArgArgGluGlnGlnLeuArgArgGluGlnGlnLeuArgArgGluGlnGlnLeuArg 403
QY 354 GAAGATGTCATCCACATGCTGAGGACAGAGAAAAACCAAGCCGAGGTTCTGAGGCACAC 413
Db 404 ArgGlu-----GlnGlnLeuArgArgGluGln----- 412
QY 414 TATGGATCTGCAGAACTTGAGAAAGTGCTTCGG-----GTCCTGCACCGAGATGCC 464
Db 413 -----GlnLeuArgGluGlnGlnLeuArgArgGluGlnGlnLeuArgArgGluGln 430
QY 465 ATCCTTGCTCAAGAGAAGTCCATAGGAGAAAGACGTCTATGAGAAACCTATCTCAGAGCTG 524
Db 431 GlnLeuArgArgGluGlnGlnLeuArgArgGluGlnGlnLeuArgArgGluGlnGluGlu 450
QY 525 GACAGACTGGAGGAAAAGCAGAAGGAGACGTTACCGCCGCATGCTAGAGCAGCTGCTGCTG 584
Db 451 GluArgHisGluGlnLysHisGluGlnGluArgArg-----GluGlnArgLeuLys 467
QY 585 GCTGAGAAGTGTACAGGCGCACCGTGTACGAGTGGAGAACGAGAGACGACACACACT 644
Db 468 ArgGluGlnGluGluArgArgAspTrpLeuLysArgGluGluGluThrGluArgHisGlu 487
QY 645 GACTACATGAACAAGACGCGACGACTTCCACCACTGCTGGAGCAGGAGCGAGAGGTTG 704
Db 488 GlnGluArgArgLysGlnGlnLeuLysArgAspGlnGluGluArgGluArgGluArgTrp 507
QY 705 AAAAAGCTCCTTGAACAAGAAAAAGCTTACCAGGCCCGCAAAGAAAAAGGAAAAACGCTAAG 764
Db 508 LeuLysLeuGluGluGluGluArgArgGluGlnGlnGluArgArgGluGlnGlnLeuArg 527
QY 765 CGGCTCAACAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCATGTTGGTG 824
Db 528 ArgGluGlnGluGluArgArgGlu---GlnArgLeuLys----- 539
QY 825 GACGAGAGGCAGATGCACATCGAGCAACTGGGCTGCAGAGTCAGAAAGTCCAGGACCTC 884
Db 540 -----ArgGlnGluGluGluArgLeu-----GlnGlnArgLeuArgSerGlu 554
QY 885 ACTCAGAAGCTGAGGAGGAGGAAGAAAACTCAAAGCGGTCACTTACAAATCCAAGGAA 944
Db 555 GlnGlnLeuArgArgGluGlnGluGluArgLeuGluGlnLeuLeuLysArgGluGluGlu 574
QY 945 GACCCGACGAAGTGTCTCAAGTTAGAAGTGGACTTCGAAACAAGGCCCTCGAGGTTTTC 1004
```

```
Db 575 LysArg-----LeuGluGlnGluArgArgGluGlnArgLeuLys 587
QY 1005 CAGGAGCACGAAGAGATGAACGCCAAATTGGCGAATCAAGAATCTCACAAACCGCAACTT 1064
Db 588 ArgGluGlnGluGluArgArgAspGlnLeuLeuLysArgGluGluGluArgArgGlnGln 607
QY 1065 CGACTCAAACTG-----GTTGGCTTATCGCAAAAGGATTGAG----- 1100
Db 608 ArgLeuLysArgGluGlnGluGluArgLeuGluGlnArgLeuLysArgGluValGlu 627
QY 1101 -----GAGCTCGAAGAGACCAATAAAAGCCTTCAGAAAGGCAGAGGAAGCTCCAGGAG 1154
Db 628 ArgLeuGluGlnGluGluArgArgAspGluArgLeuLysArgGluProGluGluGlu 647
QY 1155 CTGAGAGAGAAAAATTGCCAAAGGGGAATGTGGAAACTCCAGTCTCATGGCGGAAGTGGAG 1214
Db 648 ArgArgHisGluLeuLeuLysSerGlu----- 656
QY 1215 AGTCTGCGCAAGCGCGTGTCTTGAGATGAGGSCAAGGATGAAGAGATCACGAAGACCCGAG 1274
Db 657 -----GluGlnGluGluArgArgHisGluGlnLeuArgArgGlu 669
QY 1275 GCCCAGTCCCGGAGCTGAAGAAGAAGCTCCAAGAGGAAGAACACCACAGCAAGGAACCTT 1334
Db 670 GlnGlnGluArgGluGlnArgLeuLysArgGluGluGluGluGluArgLeuGluGln 689
QY 1335 AGACTA-----GAAGTGGAGAACTGTCAGAAAGGAGTGTCTGAGCTGGAGAAAG 1382
Db 690 ArgLeuLysArgGluHisGluGluGluArgArgGluGlnGluLeuAlaGlu---GluGlu 708
QY 1383 CTGGAGGAAGCGTTTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAG 1442
Db 709 GlnGluGlnAlaArgGluArgGileLysSerArgIleProLysTrpGlnTrpGlnLeuGlu 728
QY 1443 AAGGAGAAGAACCTA-----ACCAAAAGACCTGTCTGAACGAGCTGGAGGTGGTCAAGAGT 1496
Db 729 SerGluAlaAspAlaArgGlnSerLysValLeuLeuGluAlaProGlnAlaGlyArgAla 748
QY 1497 CGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAGGCCGAGTTAAGCCTCAAA 1556
Db 749 GluAlaProGln---GluGlnGluGluLysArgArgArgGluSerGluLeuGlnTrpGln 767
QY 1557 GATGACCTTACAAAGCTGAAGTCCTTCACTGTGATGCTGGTGGATGAGAGGAAAAATATG 1616
Db 768 Glu-----GluGluArgAlaHisArg 774
QY 1617 ATGGAGAAAAATAAAGCAAGAAAGAGAGGAAAGTGGAT-----GGGTTGAATAAAAAAC 1667
Db 775 GlnGln-----GlnGluGluGluGlnArgArgAspPheThrTrpGlnTrpGlnAlaGlu 792
QY 1668 TTTAAGGTGGAGCAGCGGAAAAAGTC---ATGGATGTACGGGAAAAAGCTAATCGAGGAAAGC 1724
Db 793 GluLysSerGluArgGlyArgGlnArgLeuSerAlaArgProProLeuArgGluGlnArg 812
QY 1725 AAGAAGCTTTTAAAAACTCAAATCTGAAATGGAGGAAAAAGGAGTACAGTCTGACAAAGGAG 1784
Db 813 GluArgGlnLeuArgAlaGluGluArgGlnGlnArgGluGlnArgPheLeuProGluGlu 832
QY 1785 AGGGAT---GAGCTGATGGGTAAACTGAGGAGCGGAAGAAAGGTCTCTGTGAACCTGAGC 1841
Db 833 GluGluLysGluGlnArgGlyArgGlnArgArgGluArgGluLysGluLeuGlnPheLeu 852
QY 1842 TGCAGTGTAGACTTACTAAAGAAGCGG----- 1874
Db 853 GluGluGluGlnLeuGlnArgArgGluArgAlaGlnGlnLeuGlnGluGluAsp 872
QY 1875 GGCATA---GAGGAGGTAGAAAGGGAATAAAACCGAGGTAGGTGCTGCAAGGGGTCTGAG 1931
Db 873 GlyLeuGlnGluAspGlnGluArgArgArgGlnGluGlnArgArgAspGlnLysTrpArg 892
QY 1932 TTCACCTGCCCGGAAGACAATAAGATCAGAGAACTAACGCTTGAAATC----- 1979
```

Db 893 TrpGlnLeuGluGluAArgLysArgArgArgHisThrLeuTyrAlaLysProAlaLeu 912
QY 1980 ---GAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTGGAGGGGACTTGATGAAG 2036
Db 913 GlnGluGlnLeuArgLysGluGlnGlnLeuLeuGlnGluGluGluGluGlnArg 932
QY 2037 ACCGAGGACGAA-----TATGACCAGTTGGAGCAGAAGTTCAGAACCGAG----- 2081
Db 933 GluGluArgGluLysArgArgArgGlnGlnGlnArgGlnTyrArgGluGluGluGln 952
QY 2082 -----CAGGATAAGGCAAACTTCCTCTCCCAGCAGCTCGAG-----GAA 2120
Db 953 LeuGlnGlnGluGluGlnLeuLeuArgGluGluArgGluLysArgArgArgGlnGlu 972
QY 2121 ATCAARACACCAATGGCCCAAGCACAAAGCCATAGAGAAGGGGAG-----GCCGTGAGC 2174
Db 973 ArgGluArgGlnTyrArgLysAspLysLysLeuGlnGlnLysGluGluGlnLeuLeuGly 992
QY 2175 CAGGAAGCCGAACCTGCGACACACAGGTTTCGGCTGGAGGAGCTAAAGTCGTGAT----- 2228
Db 993 GluGluProGluLysArgArgArg---GlnGluArgGluLysLysTyrArgGluGluGlu 1011
QY 2229 -----TTACAGCCCGAGGTGCAGGCTCTCAAGGAG----- 2258
Db 1012 GluLeuGlnGlnGluGluGlnLeuLeuArgGluGluArgGluLysArgArgArgGln 1031
QY 2259 -----AAGATCCACGAGCTGATGAACAAGGAGACCAAGGACCAAGCTGTCT 2297
Db 1032 GluTrpGluArgGlnTyrArgLysLysAspGluLeuGlnGlnGluGluGlnLeuLeu 1051
QY 2298 CAGCTCCAAGTCGACTATTCGGTCCCTTCAGCAAGA-----TTTATGGAAGAGAA 2348
Db 1052 ArgGluGluArgGluLysArgArgLeuGlnGlnArgGluArgGlnTyrArgGluGlu 1071
QY 2349 -----ACTAAGAACAAGACATGGGGAGGGAG----- 2375
Db 1072 GluLeuGlnGlnGluGluGlnLeuLeuGlyGluGluArgGluThrArgArgArgGln 1091
QY 2376 -----GTCCTCAATCTGACCAAGGAGCTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTC 2429
Db 1092 GluLeuGluArgGlnTyrArgLysGluGluGluLeuGlnGlnGlnGluGlnLeuLeu 1111
QY 2430 AGGCCGAGTGGGAACGGCCGAAGGATGGTGGACGTGCTGTGGCCTCCACTGGGGTGCAG 2489
Db 1112 ArgGluGluProGluLysArgArgArg-----GlnGlu 1122
QY 2490 ACCGAGCGGTGTGCGGGGATGCTGCGGAGGAGGAGACCCCGGCTGTGTTTCATTCGCAAA 2549
Db 1123 ArgGluArgGlnCysArgGluGluGluGluLeuGlnGlnGluGluGlnLeuLeuArg 1142
QY 2550 TCCTTCCAGGAGGAAAATCACATCATFAGTAATCTTCGACAGGTAGGCCCTGAAGAAACCC 2609
Db 1143 GluGluArgGluLysArgArgArgGlnGluLeuGluArgGlnTyrArgGluGluGlu 1162
QY 2610 ATGGAACGGTCCTCGGTCCTCGACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGG 2669
Db 1163 LeuGlnArgGlnLysArgLysGlnArgTyrArgAspGluAspGlnArgSerAspLeuLys 1182
QY 2670 AAGTCTTGGATTCCCTGGATGAGAAAAAGAGAAAAC 2705
Db 1183 TrpGlnTrpGluPro-----GluLysGluAsn 1191

RESULT 8
US-08-800-644-94
; Sequence 94, Application US/08800644
; Patent No. 5958752
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and

; TITLE OF INVENTION: Mehods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/800,644
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/056,200
; FILING DATE: 30-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-800-644-94

Alignment Scores:
Pred. No.: 3.17e-21 Length: 1898
Score: 429.00 Matches: 216
Percent Similarity: 41.20% Conservative: 168
Best Local Similarity: 23.18% Mismatches: 376
Query Match: 5.55% Indels: 172
DB: 2 Gaps: 35

US-10-788-793-1 (1-4364) x US-08-800-644-94 (1-1898)
QY 174 GAAGATGCAAAAAAGAAACAAGGCCAATCGGAAGGAGGAGGATGTCTAGGCTTCCGGAACT 233
Db 344 GluGluArgGluGlnGlnLeuArgArgGluGlnGluGluGluArgGluGlnGln 363
QY 234 ATCAAAAGGCACCTCAAAACCATCTCTGAGAAAAGTAAAGAAAAAGACTAAGAAAGTCTGTGGAG 293
Db 364 LeuArgArgGluGlnGluGluArgArgGluGlnGlnLeuArgArgGluGlnGluGlu 383
QY 294 TTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGGAGTTGCAGGCTCGA 353
Db 384 GluArgArgGluGlnGlnLeuArgArgGluGlnGlnLeuArgArgGluGlnGlnLeuArg 403
QY 354 GAAGATGTATCCACATGCTGAGGACAGAGAAAACCAAGCCCGAGGTTCTGGAGGSCACAC 413
Db 404 ArgGlu-----GlnGlnLeuArgArgGluGln----- 412
QY 414 TATGGATCTGCAGAACCTGAGAAAAGTCTTCGG-----GTCTGCACCCGAGATGCC 464
Db 413 -----GlnLeuArgArgGluGlnGlnLeuArgArgGluGlnGlnLeuArgArgGluGln 430
QY 465 ATCCTTGCTCAAGAGAAGTCCATAGGAGAAGACGCTCTATGAGAAAACCTATCTCAGAGCTG 524
Db 431 GlnLeuArgArgGluGlnGlnLeuArgArgGluGlnGlnLeuArgArgGluGlnGluGlu 450
QY 525 GACAGACTGGAGAAAAGCAGAGAGACGTCACGCCCGCATGCTAGAGCAGCTGCTGCTG 584
Db 451 GluArgHisGluGlnLysHisGluGlnGluArg-----GluGlnArgLeuLys 467

Db 1123 ArgGluArgGlnCysArgGluGluGluGluGlnGlnGluGluGlnLeuLeuArg 1142
QY 2550 TCCTTCAGGAGGAAATCATCATGAGTAATCTTCACAGGTAGGCTGAAGAACCC 2609
Db 1143 GluGluArgGluLysArgArgArgGlnGluLeuGluArgGlnTyrArgGluGluGlu 1162
QY 2610 ATGGAACGGTCCTCGGTCCTCGACAGGTATCCCCAGCAGCGAATGAGCTCACCATGAGG 2669
Db 1163 LeuGlnArgGlnLysArgLysGlnArgTyrArgAspGluAspGlnArgSerAspLeuLys 1182
QY 2670 AAGTCTTGATTCTCTGGATGAGAAAAAGAGAAAAAC 2705
Db 1183 TrpGlnTrpGluPro-----GluLysGluAsn 1191

RESULT 9

US-09-914-259-11
; Sequence 11, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-914-259-11

Alignment Scores:
Pred. No.: 6.53e-21 Length: 3878
Score: 426.00 Matches: 248
Percent Similarity: 38.77% Conservative: 194
Best Local Similarity: 21.75% Mismatches: 367
Query Match: 5.51% Indels: 331
DB: 4 Gaps: 46

US-10-788-793-1 (1-4364) x US-09-914-259-11 (1-3878)

QY 51 TTAAGGAGTCGACAACAGGTGGGAATGAGATCAGCAATCAAGGTGGAGAAAGTTTCATCT 110
Db 1647 LeuAlaGlyArgGluLysLeuCysCysGluLeuArgAsnSerSerThrGln--ThrGln 1665
QY 111 AACGGGCGATGTCCTCGGCCCAAGTCCTCCATCATCAGCAGTGATGTGGTAAAGGCCCC 170
Db 1666 AsnGly-----AsnGluAsnGlnGlyGluValGlu 1675
QY 171 TCAGAAGATGCAAAAAGAACAAAGGCCAATCGGAAGGAGGAGGATGTCATGGCTTCCGGA 230
Db 1676 GluGlnThrPheLysGluLysGluLeuAspArgLysProGluAspVal---Pro-ProG 1694
QY 231 ACTATCAAAAAGGCACCTCAAAACCATCTGGAGAAAGTGAGAAAAAGA---CTAAGAAGTCT 287
Db 1694 uIleLeuSerAsnGluArgTyrAlaLeuGlnLysAlaAsnAsnArgLeuLeuLysIleLe 1714
QY 288 GTGGAGTTATCCAAGGAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGGAGTTGCAG 347
Db 1714 uLeuGluValValLysThrThrAlaAla-ValGluGluThrIleGlyArgHisValLeuG 1734
QY 348 GCTCGAAGAAGATGTCATCCACATGCTGAGCAGACAGAAAAACCAAGCCCGAGGTTCTGGAG 407
Db 1734 ly-----IleLeuAspArgSerSerLysSerGlnSerSerAlaAs 1747
QY 408 GCACACTATGGATCTGCAGAACCTGAG---AAAGTGCTTCGGGTCTCTGCACCGGATGCC 464
Db 1747 erLeuIleTrp-ArgSerGluAlaGluAlaSerValLysSerCysValHisGluGluHis 1766

QY 465 ATCCTTGCTCAAGAGAGTCCATA-----GGAGAAGACGCTCTATGAGAAACCT 512
Db 1767 ThrArgValThrAspGluSerIleProSerTyrSerGlySerAspMetProArgAsnAsp 1786
QY 513 ATCTCAGAGCTGGACAGACTGGAGGAAAGACAGAAAGGAGACGTACCCGCGCATGCTA--- 569
Db 1787 IleAsnMetTrpSerLysValThrGluGluGlyThrGluLeuSerGlnArgLeuValArg 1806
QY 570 -----GAGCAGCTGCTGCTG----- 584
Db 1807 SerGlyPheAlaGlyThrGluIleAspProGluAsnGluGluLeuMetLeuAsnIleSer 1826
QY 585 -----GCTGAGAAG-----TGTCACAGGCGGCACCGTGTACGAG 617
Db 1827 SerArgLeuGlnAlaAlaValGluLysLeuLeuGluAlaIleSerGluThrSerSerGln 1846
QY 618 CTGGAGAACGAGAACACACACTGACTACATGACAAAGAGC-----GAC 665
Db 1847 LeuGluHisAlaLysValThrGlnThrGluLeuMetArgGluSerPheArgGlnLysGln 1866
QY 666 GACTTCACCAACCTGCTG-----GAGCAGGAGCGAGAGGTTG-----AAA 707
Db 1867 GluAlaThrGluSerLeuLysCysGlnGluLeuArgGluArgLeuHisGluGluSer 1886
QY 708 AAGCTCCTTGAACAA----- 722
Db 1887 ArgAlaArgGluGlnLeuAlaValGluLeuSerLysAlaGluGlyValIleAspGlyTyr 1906
QY 723 -----GAAAAAGCTTACCAAGCCCGCAAGAAAAAGGAAAAACGCTAAGCGGCTCAACAAA 776
Db 1907 AlaAspGluLysThrLeuPheGluArgGlnIleGlnGluLysThrAspIleIleAspArg 1926
QY 777 CTTCGAGATGAGCTTGTG-----AAGCTCAAGTCCTTCGCCCTC.815
Db 1927 LeuGluGlnGluLeuLeuCysAlaSerAsnArgLeuGlnGluLeuAlaGluGlnGln 1946
QY 816 ATGTTGTTGGACGAGAGCGCAGATGCACATCGAGCAA-----CTG 854
Db 1947 GlnIleGlnGluGluArgGluLeuLeuSerArgGlnLysGluAlaMetLysAlaGluAla 1966
QY 855 GGCTTCGACAGTCCAGAAAGTCCAGGACCTCACTCAGAAAGCTGAGGGAGGAGGAGAAAAA 914
Db 1967 GlyProValGluGlnGlnLeuLeuGlnGluThrGluLysLeuMetLysGluLysLeuGlu 1986
QY 915 CTCAAAGCGGTCACTTACAAATCCAAAGAACCCGCAACTTCGACTCAAACTGGTTGGCTTATCGCAA 1091
Db 1987 valGlnCysGlnAlaGluLysValArgAspAspLeuGlnLysGlnValLysAlaLeuGlu 2006
QY 972 GTGGACTTCGAACACACAGGCCTCGAGGTTTCCCAGGAGCAGCAGAAAGAGATGAACGCCAAA 1031
Db 2007 IleAspValGluGluGlnValSerArgPheIleGluLeuGluGlnGluLysAsnThrGlu 2026
QY 1032 TTGGCGAATCAAGAATCTCACAAACCCGCAACTTCGACTCAAACTGGTTGGCTTATCGCAA 1091
Db 2027 LeuMetAspLeuArgGlnGlnAsnGlnAlaLeuGluLysGlnLeuGluLysMetArgLys 2046
QY 1092 -----AGGATTGAGGAGCTGGAAGAGACCACCAATAAAAGCCTT 1127
Db 2047 PheLeuAspGluGlnAlaIleAspArgGluHisGluArgAspValPheGlnGlnGluIle 2066
QY 1128 CAGAAGCGCAGAGGAAAGAGCTCCAG----- 1151
Db 2067 GlnLysLeuGluGlnGlnLeuLysValProArgPheGlnProIleSerGluHisGln 2086
QY 1152 -----GAGCTGAGAGAGAAAAATTGCCAAAGGGGAATGT 1184
Db 2087 ThrArgGluValGluGlnLeuAlaAsnHisLeuLysGluLysThrAspLys-----Cys 2104
QY 1185 GGAACCTCCAGTCTCATGGCGGAAGTGGAGAGTCTGCGCAAGCGCGTGTGATGATGGAG 1244
Db 2105 -----SerGluLeuLeuLeuSerLysGluGlnLeuGlnArgAspIleGlnGlu----- 2120
QY 1245 GGCAAGGATGAAGAGATCACGAAGACCGAGGCCCGAGCTGCGGGAGCTGAAG----- 1295

Db 2121 ---ArgAsnGluGluIleGluLysLeuGluPheArgValArgGluLeuGluGlnAlaLeu 2139
QY 1296 -----AAGAAGCTCCAAGAGGAAGAACACCAC----- 1322
Db 2140 LeuValSerAlaAspThrPheGlnLysValGluAspArgLysHisPheGlyAlaValGlu 2159
QY 1323 ---AGCAAGGAACCTTAGACTAGAAAGTGGAGAAAGCTGCAGAAAGAGGATGTCTGAGCTGGAG 1379
Db 2160 AlaLysProGluLeuSerLeuGluValGln---LeuGlnAlaGluArgAspAlaIleAsp 2178
QY 1380 AAGCTGGAGGAAGCGTTACGCCGGAGTAAGTCGGAAATGCACCCAGCTCCATCTGAACCTG 1439
Db 2179 ArgLysGluLysGluIleThrAsnLeuGluGlnLeuGluGlnPheArgGluGluLeu 2198
QY 1440 GAGAAGGAGAAAGAACCTTAACCAAGACCTGCTGAACGAGCTGGAGGTGGTCAAG----- 1493
Db 2199 GluAsnLysAsnGluGluValGlnGlnLeuHisMetGlnLeuGluIleGlnLysLysGlu 2218
QY 1494 -----AGTCGAGTTAAAGAACTCGAATGCTCCGAGAGTAGACTGGAGAAGGCCGAGTTA 1547
Db 2219 SerThrArgLeuGlnGluLeuGlu---GlnGluAsnLysLeu----- 2232
QY 1548 AGCCTCAAAGATGACCTTACAAAGCTG-----AAGTCCTTCACTGTGATGCTG 1595
Db 2233 ---PheLysAspAspMetGluLysLeuGlyLeuAlaIleLysGluSerAspAlaMetSer 2251
QY 1596 GTGGATGAGAGGAAATAATATGATGGAGAAA-----ATAAAGCAAGAAGAGAGG 1643
Db 2252 ThrGlnAspGlnHisValLeuPheGlyLysPheAlaGlnIleIleGlnGluLysGluVal 2271
QY 1644 AAAGTGGATGGTGTGAATAAAACTTT---AAGTGGAGCAGGGAAGAAAGTCAATGGATGTG 1700
Db 2272 GluIleAspGlnLeuAsnGluGlnValThrLysLeuGlnGlnGlnLeuLysIleThrThr 2291
QY 1701 ACGGAAAAGCTAATCGAGGAAGCAAGAAGCTTTTAAAA----- 1739
Db 2292 AspAsnLysValIleGluGluLysAsnGluLeuIleArgAspLeuGluThrGlnIleGlu 2311
QY 1740 ---CTCAAATCTGAAATGGAG-----GAAAGGAGTACACTCTGACA 1778
Db 2312 CysLeuMetSerAspGlnGluCysValLysArgAsnArgGluGlu-----Ile 2328
QY 1779 AAGGAGAGGGATGAGCTGATGGGTAAACTGAGGACCGAA----- 1817
Db 2329 GluGlnLeuAsnGluValIleGluLysLeuGlnGlnGluLeuAlaAsnIleGlyGlnLys 2348
QY 1818 ---GAAGAAAGGTCCTGTGAACCTGAGCTGCAGTGTAGACTTACTAAAGAACGGCTTGAT 1874
Db 2349 ThrSerMetAsnAlaHisSerLeuSerGluGluAlaAspSerLeuLysHisGlnLeuAsp 2368
QY 1875 GGCATAGAGGAGGTAGAAAGGGAAATAAACCGAGGTAGTGTGTCGAAGGGTCTGAGTTC 1934
Db 2369 ValValIleAlaGluLysLeuAlaLeuGluGlnGlnValGluThrAlaAsnGluMet 2388
QY 1935 ACC-----TGCCCCGAAGACAAT---AAGATCAGAGAACTAACGCTTGAA 1976
Db 2389 ThrPheMetLysAsnValLeuLysGluThrAsnPheLysMetAsnGlnLeuThrGlnGlu 2408
QY 1977 ATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTGGTGGAGGG----- 2024
Db 2409 LeuPheSerLeuLysArgGluArgGluSerValGluLysIleGlnSerIleProGluAsn 2428
QY 2025 -----GACTTGATGAAG 2036
Db 2429 SerValAsnValAlaIleAspHisLeuSerLysAspLysProGluLeuGluValValLeu 2448
QY 2037 ACCGAGGACGAATATGACCAGTTGGAG----- 2063
Db 2449 ThrGluAspAlaLeuLysSerLeuGluAsnGlnThrTyrPheLysSerPheGluGluAsn 2468
QY 2063 ----- 2063

Db 2469 GlyLysGlySerIleIleAsnLeuGluThrArgLeuLeuGlnLeuGluSerThrValSer 2488
QY 2063 ----- 2063
Db 2489 AlaLysAspLeuGluLeuThrGlnCysTyrLysGlnIleLysAspMetGlnGlnGly 2508
QY 2064 -----CAGAAGTTCAGAAACCGAG 2081
Db 2509 GlnPheGluThrGluMetLeuGlnLysLysIleValAsnLeuGlnLysIleValGluGlu 2528
QY 2082 CAGGATAAGGCAAACTTCTCTCCAG---CAGCTCGAGGAAATCAAAACACCAAAATGGCC 2138
Db 2529 LysValAlaAlaAlaLeuValSerGlnIleGlnLeuGluAlaValGlnGluTyrAla--- 2547
QY 2139 AAGCACAAAGCCATAGAGAAAGGGGAGGCCGTGAGCCAGGAAGCCGAACCTGCACACAGG 2198
Db 2548 -----LysPheCysGlnAspAsnGlnThrIleSerSerGluProGluArgThrAsnIle 2565
QY 2199 TTTCGGCTGGAGAGGCTAAAAGTCGTGATTATCAGGCCGAGGTGCAGGCTCTCAAGGAG 2258
Db 2566 GlnAsnLeuAsnGlnLeuArgGluAspGluLeuGlySerAspIleSerAlaLeuThrLeu 2585
QY 2259 AAGATCCACGAGCTG----- 2273
Db 2586 ArgIleSerGluLeuGluSerGlnValValGluMetHisThrSerLeuIleLeuGluLys 2605
QY 2274 -----ATGAACAAGGAAGACCAGCTGTCTCAGCTC 2303
Db 2606 GluGlnValGluIleAlaGluLysAsnValLeuGluLysGluLysLysLeuLeuGluLeu 2625
QY 2304 CAAGTCGACTATTTCGGTCTCTCAGCAAAGATTATGGAAGAAGAACTAAGAACAAAGAAC 2363
Db 2626 GlnLysLeuLeuGluGlyAsnGluLysLysGlnArgGluLysGluLysLysArgSerPro 2645
QY 2364 ATGGGGAGGGAGTCTCAATCTGACCAAGGAGCTAGAGCTTTCCAAGCGGTACAGC--- 2420
Db 2646 GlnAspValGluValLeuLysThrThrThrGluLeuPheHisSerAsnGluGluSerGly 2665
QY 2421 -----CGAGCTCTCAGGCCGAGTGGGAAACGGCCGAGGATGTTGGACGTG 2465
Db 2666 PhePheAsnGluLeuGluAlaLeuArgAlaGluSerValAlaThrLys-----Ala 2682
QY 2466 CTTGTGGCTCCACTGGGGTGCAGACCCGAGCGGTGTGCGGGATGTCGGGAGGAGGAG 2525
Db 2683 GluLeuAlaSerTyrLysGluLysAlaGluLysLeuGlnGluLeuValLysGlu 2702
QY 2526 ACCCGGGCTGTTCATTTCGAAATCCTTC---CAGGAGGAAATCATCATCATGAGTAAT 2582
Db 2703 ThrAsnMetThrSerLeuGlnLysAspLeuSerGlnValArgAspHis----- 2718
QY 2583 CTTGACAGAGGTAGCCCTGAAGAAACCCATGGAACGGTCTCGTCCCTCGACAGG 2636
Db 2719 -----LeuAlaGluAlaLysGluLysLeuSerIleLeuGluLys 2731

RESULT 10

US-08-714-741-41

; Sequence 41, Application US/08714741

; Patent No. 6500613

; GENERAL INFORMATION:

; APPLICANT: Briles, David E.

; APPLICANT: McDaniel, Larry S.

; APPLICANT: Swiatlo, Edwin

; APPLICANT: Yother, Janet

; APPLICANT: Crain, Marilyn J.

; APPLICANT: Hollingshead, Susan

; APPLICANT: Tart, Rebecca

; APPLICANT: Brooks-Walter, Alexis

; TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,

; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,

; TITLE OF INVENTION: PORTIONS AND PRODUCTS

; NUMBER OF SEQUENCES: 47

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Curtis, Morris & Safford, P.C.

Db 432 GluAsnGlnValHisArgLeuGluGlnGluLeuLysGluIleAspGluSerAspSerGlu 451
QY 2028 TTGATGAAGACCGAGGACGAATATGACCAGTTGGAGCAGAAGTTCAGAAACCGAGCAG--- 2084
Db 452 AspTyrLeuLysGluGlyLeuArgAlaProLeuGlnSerLysLeuAspThrLysAla 471
QY 2085 -----GATAAGGCAAACTCCTCTCTCCAGCAGCTCGAGGAAATCAAACACCAATGGCC 2138
Db 472 LysLeuSerLysLeuGluLeuSerAspLysIleAspGluLeuAspAlaGluIleAla 491
QY 2139 AAGCACAAAGCC---ATAGAGAAAGGGAGGCCGTGAGCCAGGAAGCCGAACCTGGCAGAC 2195
Db 492 LysLeuGluValGlnLeuSerGluSerGluAspTyrAlaLysGluGly----- 507
QY 2196 AGGTTTCGGCTGGAGGAGCTAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAG 2255
Db 508 ---PheArg-----AlaProLeuGlnSerLysLeuAspAlaLysLys 520
QY 2256 GAGAAGATCCACGAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTCCAAGTCGACTAT 2315
Db 521 AlaLysLeuSerLysLeuGluLeuSerAspLysIleAspGluLeuAspAlaGluIle 540
QY 2316 TCGGTCCTTCAGCAAGATTATGGAAGAAGAACTAAGAAACAAG----- 2360
Db 541 AlaLysLeuGluAspGlnLeuLysAspAlaGluGlyAsnAsnValGluAlaTyrPhe 560
QY 2361 AACATGGGAGGGAGTGCTCTCAATCTGACCAAGGAGTAGAGCTTTCCAAAGCGCTACAGC 2420
Db 561 LysGluGlyLeuGluLysThrAlaGluLysLysAlaGluLeuGluLys----- 577
QY 2421 CGAGCTCTCAGGCCGAGTGGGAACGGCCGAAGGATGGTGGACGTGCCTGTGGCCTCCACT 2480
Db 578 -----AlaGluAlaAspLeuLysLysAlaValAspGluPro----- 589
QY 2481 GGGGTGCAGACCGGCGGTGTGCGGGGATGCTGCGGAGGAGGAGACCCCGGTGTGTTC 2540
Db 590 -----GluThrProAlaProAla 595
QY 2541 ATTCGCAAAATCCTTCAGGAGGAAATACATCATGAGTAATCTTCGACAGGTAGGCCTG 2600
Db 596 ProGlnLysAlaAlaGluGluAsnAsnValGluAspTyrPheLysGlu---GlyLeu 614
QY 2601 AAGAAACCCCATGGAAACGGTCTCGTCTCTCGACAGGTATCCCCAGCAGCGAATGAGCTC 2660
Db 615 GluLysThrIle-----AlaAlaLysLysAla 623
QY 2661 ACCATGAGGAAGTCTTGATTCCTTGATGAGAAAGAGAAACGGTCTCTCCACTCCG 2720
Db 624 GluLeuGluLysThrGluAlaAsp---LeuLysLysAlaValAsnGluProGluLysPro 642
QY 2721 -----CAGGAGAAAGGGCCCGCCAGGCCAAACCCAGGTGCAGGGCACCCCGGGGAGCTG 2771
Db 643 AlaProAlaProGluProAlaProAlaProGluLysProAlaGluLysPro----- 659
QY 2772 GTCCTAGCACCAAGCAGGGCCAGCCCTACATCCGTGTGACACCAGATCATGAGAAC 2831
Db 660 -----AlaProAlaProGluLysPro----- 666
QY 2832 AGCACTGCCACCCTGGAGATCACAAGCCCCACATCTGAAGAGTTTTTCTCTAGTACCACC 2891
Db 667 -----AlaProAlaProGluLys----- 672
QY 2892 GTCATTCTACCTTAGGCAACACAGAAACCAAGATAACCATATTATCCATCACCAATGTC 2951
Db 673 -----ProAlaProAlaProGluLysPro---AlaProAlaThrProAlaProGluAla 689
QY 2952 ATGTCGCAAAAGCCCCAAAAGTGCAGATCCTACTCTCGGCCCCAGAACGAGCCATGTCCCCT 3011
Db 690 ProAlaGluGlnProLys-----ProAlaProAlaProGlnProAlaProAlaPro 706
QY 3012 GTCACGATTACTATTTCAGAGAGAAG-----AGCCCGGAAGGTGGAAGG 3059
Db 707 LysProGluLysProAlaGluGlnProLysProGluLysThrAspAspGlnAlaGlu 726

QY 3060 AGCGCCTTTGCCGACAGGCCTGCATCCCCCATCCAAATCATGACGGTGTCAACATCTGCA 3119
Db 727 GluAspTyrAlaArgArgProGluLysPro-----AlaPro 738
QY 3120 GCTCCCACTGAATCGCTGTCTCTCTCTGAATCT-----CAGGAAGTG 3161
Db 739 AlaProGluLysProAlaProThrProGluThrProLysThrGlyTyrLysGlnGluAsn 758
QY 3162 CCTATGGGA----- 3170
Db 759 GlyMetTyrPheTyrAsnThrAspGlySerMetAlaThrGlyTyrSerGluGluGlu 778
QY 3171 ---AGACTATCTCAAGTCAACCCCGGAAAAACAACAACTGTTCCAGCCCCCGTCGGGAAG 3227
Db 779 TyrAsnArgLeuThrGlnGlnProProLysAlaGluLysProAlaProAlaProLys 798
QY 3228 -----TACAACTCCAATGCTAATATCATC 3251
Db 799 ThrGlyTyrLysGlnGluAsnGlyMetTyrPheTyrAsnThrAspGlySerLeu--- 817
QY 3252 ACCACGGAAGACATAAAATTCACATTCACCTGGGTCTCAGTTTAAGCGATCTCCTGGG 3311
Db 818 -----GlnAsnAsnGlySerTyrTyrTyrLeuAsnSerAsnGlyAlaMetAlaThrGly 835

RESULT 11

US-09-310-187A-1
; Sequence 1, Application US/09310187A
; Patent No. 6358751
; GENERAL INFORMATION:
; APPLICANT: Benichou, Gilles
; APPLICANT: Fedoseyeva, Eugenia
; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
; TITLE OF INVENTION: Graft Rejection
; FILE REFERENCE: UCSF-090
; CURRENT APPLICATION NUMBER: US/09/310,187A
; CURRENT FILING DATE: 1999-05-12
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-310-187A-1

Alignment Scores:
Pred. No.: 6.93e-20 Length: 1939
Score: 409.50 Matches: 200
Percent Similarity: 40.23% Conservative: 181
Best Local Similarity: 21.12% Mismatches: 297
Query Match: 5.30% Indels: 269
DB: 4 Gaps: 37

US-10-788-793-1 (1-4364) x US-09-310-187A-1 (1-1939)

QY 177 GATGCAAAAAAGAACCAAGCCCAATCGGAAG-----GAGGAGGATGTATGGCTTCC 227
Db 918 GluAlaLysValLysGluMetAsnGluArgLeuGluAspGluGluMetAsnAlaGlu 937
QY 228 GGAACATATCAAAAAGGCACCTCAAACCATCTCGAGAAAGTGAGAAAAAGACTAAGAAGTCT 287
Db 938 LeuThrAlaLysLys-----ArgLysLeuGluAspGluCys 949
QY 288 GTGGAGTTATCCAAG-----GAGGACCTCATCCAGCTCCTGAGTATCATGGAAGGGGAG 341
Db 950 SerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLysGlu 969
QY 342 TTGCAGGCTCGAGAAAGATGTCTATCCACATGCTGAGGACAGAG----- 383
Db 970 LysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAspGlu 989
QY 384 -----AAAACCAAGCCCGAGGTCTTGAGGCGCACACTATGATCTGCA 425

Db 990 IleIleAlaLysLeuThrLysGluLysLysAlaLeuGlnGluAlaHisGlnGlnAlaLeu 1009
QY 426 GAACCTGAGAAAGTCTTCGGGTCTGCACCGAGATGCCATCCTTGCTCAAGAGAGTCC 485
Db 1010 AspAspLeuGlnVal-----GluGluAspLysValAsnSerLysSer 1025
QY 486 ATAGGAGAGACGTCTATGAGAAACCTATCTCAGAGCTGGAC---AGACTGGAGGAAAG 542
Db 1026 -----LysValLysLeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGlu 1043
QY 543 CAGAAG-----GAGACGTACCGCCGCATGCTAGAGCAGCTGCTGGCT 587
Db 1044 LysLysValArgMetAspLeuGluArgAlaLysArgLysLeuGluGlyAspLeu----- 1061
QY 588 GAGAAAGTGTACAGCGGCACCGCTGTACGAGCTGGAGAACGAGAGACACAGCAGCTGAC 647
Db 1062 ---LysLeuThrGlnGluSerIleMetAspLeuGluAsnAspLysLeuGlnLeuGluGlu 1080
QY 648 TACATGAACAAGAGCGACGACTTCACCAACCTGCTGGACGAGGAGCGAGAGGTTGAAA 707
Db 1081 LysLeuLysLysLysGluPheAspIleAsn-----GlnGlnAsnSerLysIleGlu 1097
QY 708 AAGCTCTCTGAACAAGAAAGAGCTTACCAAGCCCGCAAGAA---AAGGAAACCGCTAAG 764
Db 1098 Asp-----GluGlnAlaLeuAlaLeuGlnLeuLysLysLysLeuLysGluAsnGlnAla 1115
QY 765 CGGCTCAACAACACTTCGAGATGAGCTT-----GTG 794
Db 1116 ArgIleGluGluLeuGluGluGluAlaGluArgThrAlaArgAlaLysValGlu 1135
QY 795 AAGCTCAAGTCTCTTCGCCCTCATGTTGGTGGACGAGAGCGAGATGCACATCGAGCAACTG 854
Db 1136 LysLeuArgSerAspLeuSerArgGluLeuGluGluIleSerGluArgLeuGluGluAla 1155
QY 855 GGCCTCGAGAGTCAGAAAGTCCAGGACCTCACTCAGAGCTGAGGGAGGAGGAAGAAAA 914
Db 1156 GlyGlyAlaThrSerValGlnIleGluMetAsnLysLysArgGluAlaGluPheGlnLys 1175
QY 915 CTCAAGCGGTCACTTACAAATCCAAGGAAGACCGCCAGAGCTGCTCAAGTTAGAGTG 974
Db 1176 Met-----ArgArgAspLeuGluGluAlaThrLeu 1185
QY 975 GACTTCGAACACAAGGCTCGAGGTTTTCCAGGAGCAGCAAGAGATGAACGCCAAATTG 1034
Db 1186 GlnHisGluAlaThrAlaAlaAlaLeuArgLysLysHisAlaAspSerValAlaGluLeu 1205
QY 1035 GCGAATCAA-----GAATCTCACAAACCCGGCAA 1061
Db 1206 GlyGluGlnIleAspAsnLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGlu 1225
QY 1062 CTTCCGACTCAAACTGGTGGCTTATCGCAAGGATTGAGGAGCTGGAAAGAGACCAATAAA 1121
Db 1226 PheLysLeuGluLeuAspAspValThrSerAsnMetGluGlnIleIleLysAlaLysAla 1245
QY 1122 AGCCTTCAGAAGGCA-----GAGGAAGAGCTCCAGGAGCTGAGAGAGAAAAATT 1169
Db 1246 AsnLeuGluLysValSerArgThrLeuGluAspGlnAlaAsnGluTyrArgValLysLeu 1265
QY 1170 -----GCCAAAGGGGAATGTGGA 1187
Db 1266 GluGluAlaGlnArgSerLeuAsnAspPheThrThrGlnArgAlaLysLeuGlnThrGlu 1285
QY 1188 AACTCCAGTCTCATGCGGGAAGTGGAGAGTCTGCGC----- 1223
Db 1286 AsnGlyGluLeuAlaArgGlnLeuGluGluLysGluAlaLeuIleSerGlnLeuThrArg 1305
QY 1224 -----AAGCGCTGCTTGATGAGAGGC 1247
Db 1306 GlyLysLeuSerTyrThrGlnGlnMetGluAspLeuLysArgGlnLeuGluGluGly 1325
QY 1248 AAGGATGAAGAG-----ATCAGCAAGACCGAGGCCCTGCTCCGGGAGCTG 1292
Db 1326 LysAlaLysAsnAlaLeuAlaHisAlaLeuGlnSerAlaArgHisAspCysAspLeuLeu 1345

QY 1293 AAGAAAGAGCTCCAAGAGGAAGAACACCACAGCAAGGAACCTTAGACTA----- 1340
Db 1346 ArgGluGlnTyrGluGluGluThrGluAlaLysAlaGluLeuGlnArgValLeuSerLys 1365
QY 1340 ----- 1340
Db 1366 AlaAsnSerGluValAlaGlnTrpArgThrLysTyrGluThrAspAlaIleGlnArgThr 1385
QY 1341 --GAAGTGGAGAGCTGCAGAAAGAGGATGCTCTGAG---CTGGAGAAAGCTGGAGGAAGCG 1394
Db 1386 GluGluLeuGluGluAlaLysLysLysLeuAlaGlnArgLeuGlnAspAlaGluGluAla 1405
QY 1395 TTCAGCCGGAGTAAGTCGGAATGCACCAGCTCCATCTGAACCTGGAGAGGAGGAAGAAC 1454
Db 1406 ValGluAlaValAsnAlaLysCysSerSerLeuGluLysThrLysHisArgLeuGlnAsn 1425
QY 1455 CTAACCAAGACCTGCTGAACGAGCTGGAG----- 1484
Db 1426 GluIleGluAspLeuMetValAspValGluArgSerAsnAlaAlaAlaLeuAsp 1445
QY 1485 -----GTGGTCAAGAGTCGAGTTAAAGAACTCGAATGCTCC 1520
Db 1446 LysLysGlnArgAsnPheAspLysIleLeuAlaGluTrpLysGlnLysTyrGluGluSer 1465
QY 1521 GAGAGTAGACTGGAG-----AAGCCGAGTTAAGCCTCAAAGATGACCTTACAAAG 1571
Db 1466 GlnSerGluLeuGluSerSerGlnLysGluAlaArgSerLeuSerThrGluLeuPheLys 1485
QY 1572 CTGAAGTCTTCACTGCTGATGCTGGTGGATGAGAGGAAAAAATATGATGGAGAAAAATAAAG 1631
Db 1486 LeuLysAsn-----AlaTyrGluGluSerLeuGluHisLeuGluThrPheLys 1501
QY 1632 CAAGAAAGAGAGAACTGGATGGTTGAATAAAACTTTAAGGTGGAGCAGGGGAAAAGTC 1691
Db 1502 ArgGlu-----AsnLysAsnLeu-----GlnGluGluIle 1511
QY 1692 ATGGATGTGACGGAAGAAAGCTAATCGAGGAAAGCAAGAGCTT----- 1733
Db 1512 SerAspLeuThrGluGlnLeuGlyGluGlyLysAsnValHisGluLeuGluLysVal 1531
QY 1734 -----TTAAAACTCAAACTGAAATCGAGGAAAGCAAGAGCTT----- 1769
Db 1532 ArgLysGlnLeuGluValGluLysLeuGluLeuGlnSerAlaLeuGluGluAlaGluAla 1551
QY 1770 AGTCTGACAAAGGAGAGGGATGAGCTGATG-----GGTAAACTGAGG 1811
Db 1552 SerLeuGluHisGluGluGlyLysIleLeuArgAlaGlnLeuGluPheAsnGlnIleLys 1571
QY 1812 AGCGAAGAAAGAAAGTCTCTGT----- 1832
Db 1572 AlaGluIleGluArgLysLeuAlaGluLysAspGluGluMetGluGlnAlaLysArgAsn 1591
QY 1833 -----GAACTGAGCTGCAGTGTAGACTTA----- 1856
Db 1592 HisGlnArgValValAspSerLeuGlnThrSerLeuAspAlaGluThrArgSerArgAsn 1611
QY 1857 -----CTAAAGAAAGCGGCTTGATGGC---ATAGAGGAGGTAGAAAGGAAATA 1901
Db 1612 GluValLeuArgValLysLysLysMetGluGlyAspLeuAsnGluMetGluIleGlnLeu 1631
QY 1902 AACCGAGGTAGTCTGTCGAAGGGGTCTGAGTTTCACCTGCCGGAAGACAATAAGATCAGA 1961
Db 1632 SerHisAla-----AsnArgMetAlaAla 1639
QY 1962 GAACCTAACGCTTGAATCGAGAGACTGAAGAAACCGGCTCCAGCAGTTGGAGGTGGGAG 2021
Db 1640 GluAlaGlnLysGlnValLysSerLeuGlnSerLeuLysAspThrGlnIleGlnLeu 1659
QY 2022 GGGGACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAGCAGAGATTTCAGAACCGAG 2081
Db 1660 AspAspAlaValArgAlaAsnAspAsp-----LeuLysGluAsnIleAlaIleVal 1676

```
QY 2082 CAGGATAAGGCAAACTTCCTCTCCAGCAGCTCGAGGAAATCAAA----- 2126
   ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||
Db 1677 GluArgArgAsnLeuLeuGlnAlaGluLeuGluGluLeuArgAlaValGluGln 1696
   ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||
QY 2127 -----CACCAATGGCCAAAGCACAAAGCCATAGAGAAAGGGAGGCCGTG--- 2171
   ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||
Db 1697 ThrGluArgSerArgLysLeuAlaGluGlnGluLeuIleGluThrSerGluArgValGln 1716
   ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||
QY 2172 -----AGCCAGGAAGCCGAACTGCGACACACAGGTTTCGGCTGGAGGAGGCTAAAAAGT 2222
   |||||  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||
Db 1717 LeuLeuHisSerGlnAsnThrSerLeuIleAsnGlnLysLysLysMetGluSerAspLeu 1736
   |||||  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||
QY 2223 CGTGATTACAGGCCGAGGTGCAGGCTCTCAAGGAGAAGATCCACGAGCTGATGAACAAG 2282
   |||||  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||
Db 1737 ThrGlnLeuGlnSerGluValGlu-----GluAlaValGlnGluCysArgAsnAla 1753
   |||||  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||
QY 2283 GAAGACCAGCTGTCTCAGCTCCAAGTCGACTATTTCGGTCCTTCAGCAAAGATTATGGAA 2342
   |||||  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||
Db 1754 GluGluLysAlaLysLysAlaIleThrAspAlaAlaMetMetAlaGluGluLeuLysLys 1773
   |||||  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||
QY 2343 GAA-----GAAACTAAGAACAAACATGGGGAGGGAGGTCCTC 2381
   |||  |||  |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
Db 1774 GluGlnAspThrSerAlaHisLeuGluArgMetLysLysAsnMetGluGlnThrIleLys 1793
   |||  |||  |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
QY 2382 AATCTGACCAAGGAGCTAGAG 2402
   ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||
Db 1794 AspLeuGlnHisArgLeuAsp 1800
   ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||

RESULT 12
US-08-466-390-4
; Sequence 4, Application US/08466390
; Patent No. 5686562
; GENERAL INFORMATION:
; APPLICANT: TOUTKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,390
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-466-390-4
```

Alignment Scores: 7.12e-20 Length: 2101
Pred. No.: 409.50 Matches: 293
Score: 36.09% Conservative: 249
Percent Similarity:

```
Best Local Similarity: 19.51% Mismatches: 583
Query Match: 5.30% Indels: 377
DB: 1 Gaps: 59

US-10-788-793-1 (1-4364) x US-08-466-390-4 (1-2101)

QY 51 TTAAGGAGTCGACAAACAGGTGGGAATGAGATCACGAAATCAAGGTGGAGAAAGTTTCATCT 110
   |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 680 LeuArgSerGluGlnGlnLysAlaThrGluLysGluArgValAlaGlnGluLysAspGln 699
   |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 111 ---AACGGGCATGTCCTCTGCCCCAAGTCCTCCATCATCAGCAGTGATGGTGGTAAGGGC 167.
   ::::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 700 LeuGlnGluGlnLeuGlnAlaLeuLysGluSerLeuLysValThr-----LysGly 716
   ::::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 168 CCTCAGAAGATGCAAAAAGAAACAAGGCCCAATCGGAAGGAGGATGTCATCGCTTCC 227
   |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 717 SerLeuGluGluLysArgArgAlaAlaAspAlaLeuGluGlnGlnArgCysIle 736
   |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 228 GGAACCTATCAAAAGGCACCTCAAACCATCTCGAGAAAGTGAAGAAAGACTAAGAAAGTCT 287
   ::::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 737 SerGluLeuLysAlaGluThrArgSerLeuValGluGlnHisLysArgGluArgLysGlu 756
   ::::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 288 GTGAGATTATCCAAGGAGGAC-----CTCATCCAGCTCCTGAGT 326
   ::::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 757 LeuGluGluGluArgAlaGlyArgLysGlyLeuGluAlaArgLeuLeuGlnLeuGlyGlu 776
   ::::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 327 ATCATGGAAGGGGAGTTGCAGGCT---CGAGAAGATGTATCCACATGCTGAGGACAGAG 383
   ::::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 777 AlaHisGlnAlaGluThrGluValLeuArgArgGluLeuAlaGluAlaMetAlaAlaGln 796
   ::::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 384 AAAACCAAG-----CCCGAGGTTCTG----- 404
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 797 HisThrAlaGluSerGluCysGluGlnLeuValLysGluValAlaAlaTrpArgAspGly 816
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 405 -----GAGGCACACTATGATCTGCAGAACCTGAGAAAGTGCTT 443
   |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 817 TyrGluAspSerGlnGlnGluGluAlaGlnTyrGlyAlaMetPheGlnGluGlnLeuMet 836
   |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 444 CGGTC-----CTGCACCGAGATGCCATCCTTGTCTCAAGAGAGTCC 485
   ::::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 837 ThrLeuLysGluGluCysGluLysAlaArgGlnGluLeuGlnGluAlaLysGluLysVal 856
   ::::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 486 ATAGGA----- 491
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 857 AlaGlyIleGluSerHisSerGluLeuGlnIleSerArgGlnGlnAsnLysLeuAlaGlu 876
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 492 -----GAGACGCTCTATGAGAAACCTATC----- 515
   ::::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 877 LeuHisAlaAsnLeuAlaArgAlaLeuGlnGlnValGlnGluLysGluValArgAlaGln 896
   ::::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 516 -----TCAGAGCTGGACAGACTGGAGGAAAAGCAGAGGAGACGTACCGCCGCATG 566
   ::::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 897 LysLeuAlaAspAspLeuSerThrLeuGlnGluLysMetAlaAlaThrSerLysGluVal 916
   ::::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 567 -----CTAGAGCAGCTGCTG-----CTGGCTGAGAAAGTGTACAGGCGC 605
   |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 917 AlaArgLeuGluThrLeuValArgLysAlaGlyGluGlnGlnGluThrAlaSerArgGlu 936
   |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 606 ACCGTGTACGAGCTGGAGAACGAGAACGAGAACACACACTGACTACATG----- 653
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 937 LeuValLysGluProAlaArgAlaGlyAspArgGlnProGluTrpLeuGluGlnGln 956
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 654 -----AACAAAGACGACGAC 668
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 957 GlyArgGlnPheCysSerThrGlnAlaAlaLeuGlnAlaMetGluArgGluAlaGluGln 976
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 669 TTCACCAACCTGCTGGAG----- 686
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 977 MetGlyAsnGluLeuGluArgLeuArgAlaAlaLeuMetGluSerGlnGlyGlnGln 996
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 687 -----CAGGAGCGGAGAGGTTG 704
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 997 GluGluArgGlyGlnGlnGluArgGluValAlaArgLeuThrGlnGluArgGlyArgAla 1016
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

Qy 705 AAAAGTCTCTTGAACAAAGAAAAAGCTTACCAAGCCCGCAAGAAAAAGGAAAAACGCTAAG 764
Db : : : : :
GlnAlaAspLeuAlaLeuGluLysAlaAlaArgAlaGluLeuGluMetArgLeuGlnAsn 1036
Qy 765 CGGCTCAACAAACTTCGAGATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCATGTTGGTG 824
Db : : : : :
AlaLeuAsnGluGlnArgValGluPheAlaThrLeuGlnGluAlaLeuAlaHisAlaLeu 1056
Qy 825 GACGAGGCGAGATGCACATCGAGCAACTG-----GGCCTGCAGAGTCAGAAA 872
Db : : : : :
ThrGluLysGluGlyLysAspGlnGluLeuAlaLysLeuArgGlyLeuGluAlaAlaGln 1076
Qy 873 GTCCAGGACCTCACTCAGAAGCTGAGGAGGAGGAAGAAAACTCAAAGCGGTCACTTAC 932
Db : : : : :
IleLysGluLeu--GluGluLeuArgGlnThrValLysGlnLeuLysGluGlnLeuAla 1095
Qy 933 AAATCCAGGAAGAC----- 947
Db : : : : :
LysLysGluLysGluHisAlaSerGlySerGlyAlaGlnSerGluAlaAlaGlyArgThr 1115
Qy 948 -----CGCCAGAAGCTGCTCAAGTTAGAAAGTGGAC 977
Db : : : : :
GluProThrGlyProLysLeuGluAlaLeuArgAlaGluValSerLysLeuGluGlnGln 1135
Qy 978 TTCGAACACAAGGCCTCGAGGTTTCCCGAGGAGCACGAAGAGATGAACCCAAATTGGCG 1037
Db : : : : :
CysGlnLysGlnGlnGluGlnAlaAspSerLeuGluArgSerLeuGluAlaGluArgAla 1155
Qy 1038 AATCAAGAATCTCACAAACCGGCAACTTCGACTCAAACCTGGTGGCTTATCGCAAAAGGATT 1097
Db : : : : :
SerArgAlaGluArgAspSerAlaLeuGluThrLeuGlnGlyGlnLeuGluGluLysAla 1175
Qy 1098 GAGGAGCTGGAAGAGACCAATAAAAGCCCTTCAGAAAGGCAGGAAGAGCTCCAGGAGCTG 1157
Db : : : : :
GlnGluLeuGlyHisSerGlnSerAlaLeuAlaSerAlaGlnArgGluLeuAlaAlaPhe 1195
Qy 1158 AGAGAGAAA-----ATTGCCAAA 1175
Db : : : : :
ArgThrLysValGlnAspHisSerLysAlaGluAspGluTrpLysAlaGlnValAlaArg 1215
Qy 1176 GGG-----GAATCTGGAAC-----TCCAGTCTCATGGCGAAGTGGAG 1214
Db : : : : :
GlyArgGlnGluAlaGluArgLysAsnSerLeuIleSerSerLeuGluGluValSer 1235
Qy 1215 AGTCTGCGAAGCGCGTGTGATGGAGGGCAAGGATGAAGAGATC----- 1262
Db : : : : :
IleLeuAsnArgGlnValLeuGluLysGluGlyGluSerLysGluLeuLysArgLeuVal 1255
Qy 1263 -----ACGAAGCCGAGGCCAGTGC----- 1283
Db : : : : :
MetAlaGluSerGluLysSerGlnLysLeuGluGluSerCysAlaCysArgGlnArg 1275
Qy 1283 ----- 1283
Db : : : : :
GlnProAlaThrValProGluLeuGlnAsnAlaAlaLeuLeuCysGlyArgArgCysArg 1295
Qy 1284 -----CGGAGCTGAAGAAGAAAGCTCCAAGAGGAAGACACCAAGCAAGGAACCTT 1334
Db : : : : :
AlaSerGlyArgGluAlaGluLysGlnArgValAlaSerGluAsnLeuArgGlnGluLeu 1315
Qy 1335 AGACTAGAACTGGAGAAGCTGCAGAAAGAGGATGTCTGAGCTGGAGAAGCTGGAGGAAGCG 1394
Db : : : : :
ThrSerGlnAlaGluArgAlaGluGluLeuGlyGlnGluLeuLysAlaTrpGlnGluLys 1335
Qy 1395 TTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAAGAGAAGAAC 1454
Db : : : : :
PhePheGlnLysGluGlnAlaLeuSerThrLeuGlnLeu-----GluHisThr 1351
Qy 1455 CTAACCAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAA 1514
Db : : : : :
SerThrGlnAlaLeuValSerGluLeuLeuProAlaLysHisLeuCysGlnGlnLeuGln 1371

Qy 1515 TGCTCCGAGAGTAGACTGGAGAAAGGCCGAGTTAAGCCTCAAAAGATGACCTTACAAAGCTG 1574
Db : : : : :
AlaGluGlnAlaAlaAlaGluLysArgHis----- 1381
Qy 1575 AAGTCCTTCACTGTGATGCTGCTGGTGTGATGAGAGGAAAAATATGATGGAGAAAAATAAGCAA 1634
Db : : : : :
ArgGluGluLeuGluGln 1387
Qy 1635 GAAGAGAGGAAAGTGGATGGGTTGAATAAAAACTTT-----AAGGTGGAGCAGGGA 1685
Db : : : : :
SerLysGlnAlaAlaAlaGlyGlyLeuArgAlaGluLeuLeuArgAlaGlnArgGluLeuGly 1407
Qy 1686 AAAGTCATGGATGTGACGGAAGAAAGCTAATCCGAGGAAAGCAAGAAAGCTTTTAAAACTCAA 1745
Db : : : : :
GluLeuIleProLeuArgGlnLysValAlaGluGlnGluArgThrAlaGlnGlnLeuArg 1427
Qy 1746 TCTGAATGGAGGAA-----AAGGAGTACAGTCTGACAAAGGAGAGGGATGAGCTGATG 1799
Db : : : : :
AlaGluLysAlaSerTyrAlaGluGlnLeuSerMetLeuLysLysAlaHisGlyLeuLeu 1447
Qy 1800 GGTAAACTGAGGAGCGAAGAAAGGTCCTGTGAACTGAGCTGCAGTGTAGACTTACTA 1859
Db : : : : :
AlaGluGluAsnArgGly-----Leu 1454
Qy 1860 AAGAAGCGGCTTGAT---GGCATAGAGAGGTAGAAAGGGAATAAAACCGAGGTAGGTGCG 1916
Db : : : : :
GlyGluArgAlaAsnLeuGlyArgGlnPheLeuGluValGluLeuAspGlnAlaArgGlu 1474
Qy 1917 TGCAAGGGGTCTGAGTTCACC-----TGCCCCGGAAGACAATAAGATCAGAGAACTA 1967
Db : : : : :
LysTyrValGlnGluLeuAlaAlaValArgAlaAspAlaGluThrArgLeuAlaGluVal 1494
Qy 1968 ACCTTTGAAATCGAGAGACTGAAGAAAACGGCTCCAGCAGTTCGAGGTG---GTGGAGGGG 2024
Db : : : : :
GlnArgGluAlaGlnSerThrAlaArgGluLeuGluValMetThrAlaLysTyrGluGly 1514
Qy 2025 GACTTGATGAAGACCGGAGCAGCAATATGACCAGTTGGAGCAGAAGTTTCAGAACCCGAGCAG 2084
Db : : : : :
AlaLysValLysValLeuGluGlu-----ArgGlnArgPheGlnGluGluArg 1530
Qy 2085 GATAAGGCCAAACTTCCTCTCCCAGCAGCTCGAGGAAATCAAAACACCAAAATGGCCAAAGCAC 2144
Db : : : : :
GlnLys-----LeuThrAlaGlnValGluGluLeuSerLysLysLeuAlaAspSer 1547
Qy 2145 AAAGCCATAGAAAGGGGAGGCGCTGAGCCAGGAAGCCGAACTCGACACAGGTTTCGG 2204
Db : : : : :
AspGlnAlaSerLysValGlnGlnGln-----LysLeuLys 1559
Qy 2205 CTGGAGGAGCTAAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAGGAGAGATC 2264
Db : : : : :
AlaValGlnAlaGlnGlyGlyGluSerGlnGlnGluAlaGlnArgPheGlnAlaGlnLeu 1579
Qy 2265 CACGAGCTGATGAACAAGGAAGACCAGCTGTCTCAGCTCCAAGTCCGACTATTTCGGTCCCTT 2324
Db : : : : :
AsnGluLeuGlnAlaGlnLeuSerGlnLysGluGlnAlaAlaGluHisTyrLysLeuGln 1599
Qy 2325 CAGCAAAAGATTTATGGAAGAAAGAACTAAGAAACAAGAACATGGGGAGGAGTCTCTCAAT 2384
Db : : : : :
MetGluLysAlaLysThrHisTyrAspAlaLysLysGlnGlnAsnGlnGluLeuGlnGlu 1619
Qy 2385 CTGACCAAGGAGCTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAAC 2444
Db : : : : :
GlnLeuArgSerLeuGluGlnLeuGlnLysGluAsnLysGluLeuArgAlaGluAla--- 1638
Qy 2445 GGCCGAAGGATGGTGGACGTGCTGTGGCTTCACCTGGGGTGCAGACC-----GAG 2495
Db : : : : :
GluArgLeuGlyHisGluLeuGlnGlnAlaGlyLeuLysThrLysGluAlaGlu 1656
Qy 2496 GCGGTGTGCGGGGATGCTGCGGAGGAGGAGACCCCGGCTGTGTTTCATTTCGCAATCCTTC 2555
Db : : : : :
GlnThrCysArgHisLeuThrAlaGln-----ValArgSerLeuGlu 1670
Qy 2556 CAGGAGGAAAAATCACATCATGAGTAATCTTCGACAGGTAGGCCTGAAGAAACCCATCGAA 2615

Db 1671 AlaGlnValAlaHisAlaAspGlnGlnLeuArgAspLeuGly---LysPheGlnValAla 1689
QY 2616 CGGTCCTCGGTCCTCGACAGGTATCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCT 2675
Db 1690 ThrAspAlaLeuLysSerArgGluProGlnAlaLysProGlnLeuAspLeuSerIleAsp 1709
QY 2676 TGGATTCTTGGATGAGAAAAAGAGAAAACGGTCTCTCCACTCCGCGAGGAGAAAGGCC 2735
Db 1710 SerLeuAsp---LeuSerCysGluGluGlyThrProLeuSerIleThrSerLysLeuPro 1728
QY 2736 AGCCCAAC---CAGGTCAGGGCACCCCGGGAGCTGGTC----- 2774
Db 1729 ArgThrGlnProAspGlyThrSerValProGlyGluProAlaSerProIleSerGlnArg 1748
QY 2775 CTAGCACCAAAGAGGGCCAGCCCTACACATCCGTGTGACACCATCATGAGAACAGC 2834
Db 1749 LeuProProLysValGluSerLeuGluSerLeuTyrPheThrProIleProAlaArgSer 1768
QY 2835 ACTGCCACCCTGGAGATCACAAAGCCCCACATCTGAAGAGTTTTC----- 2879
Db 1769 GlnAlaProLeuGluSerSerLeuAspSerLeuGlyAspValPheLeuAspSerGlyArg 1788
QY 2880 -----TCTAGTACCACCGTCATTCTACCTTAGGCAACCCAGAAA 2918
Db 1789 LysThrArgSerAlaArgArgThrThrGlnIleIleAsnIleThrMetThrLysLys 1808
QY 2919 CCAAGAATAACCATATTCCATCACCCCAATGTGTCGCAAAAGCCCAAAAGTGCAGAT 2978
Db 1809 LeuAspValGluGluProAspSerAlaAsnSerSerPheTyrSerThrArgSerAlaPro 1828
QY 2979 CCTACTCTCGGCCCCAGAACGAGCCATGTCCCTGTCTCAGATTACTATTCTCAGAGAG 3038
Db 1829 AlaSerGlnAlaSerLeuArgAlaThrSer-----SerThrGlnSerLeuAlaArgLeu 1846
QY 3039 AAGAGCCCGGAAGGTGAAGGAGCGCCTTTGCCAGACAGGCTGCA---TCGCCCATCCAA 3095
Db 1847 GlySerProAspTyrGlyAsnSerAlaLeuLeuSerLeuProGlyTyrArgPro----- 1864
QY 3096 ATCATGACGGTGTCAACATCTGCAGCTCCCACTCCCAATCGTGTCTCTCTGAATCTCAG 3155
Db 1865 -----ThrThrArgSerSerAlaArgArgSerGlnAlaGlyVal-----SerSer 1879
QY 3156 GAAGTGCCTATGGAAGGACTATCCTCAAAGTCAACCCCGGAAAAACAACACTGTTCCAGCC 3215
Db 1880 GlyAlaProProGlyArgAsnSerPheTyrMet---GlyThrCysGlnAspGluProGlu 1898
QY 3216 CCCGTGCGGAAGTACAACCTCCAATGTCTAATATCATCACCGGAAGACATAAAATT--- 3272
Db 1899 GlnLeuAspAspTrpAsnArgIleAlaGluLeu-----GlnGlnArgAsnArgValCys 1916
QY 3273 -----CACATTACCTGGGTTCTCAGTTTAAGCGATCTCCTGGCCTCGGCTGAAGGC 3326
Db 1917 ProProHisLeuLysThrCysTyrProLeuGluSerArgProSerLeuSerLeuGlyThr 1936
QY 3327 GTGAGC-----CCAGTTATCACCGTCCGG----- 3350
Db 1937 IleThrAspGluGluMetLysThrGlyAspProGlnGluThrLeuArgArgAlaSerMet 1956
QY 3351 ---CCTGTCAACGTGACAGCGGAGAGGAGGTTTCTACAGGCACAGTCTCGCTCTCCC 3407
Db 1957 GlnProIleGlnIle---AlaGlu-----GlyThrGlyIleThrThrArgGlnGln 1972
QY 3408 AGGAACCACTCTCT-----TCAAGACCCGGTGTGTAGCAAAAGTG 3446
Db 1973 ArgLysArgValSerLeuGluProHisGlnGlyProGlyThrProGluSerLysAla 1992
QY 3447 ACCAGCACTATA-----ACTATAACCCCG-----GTC 3473
Db 1993 ThrSerCysPheProArgProMetThrProArgAspArgHisGluGlyArgLysGlnSer 2012
QY 3474 ACAACGTATCCACAGGAGAACCCCAATCAGTGTCTCAGGACAAGATGGGTCTATCTCAGCGG 3533

Db 2013 ThrThrGluAlaGlnLysLysAlaAlaProAlaSerThrLysGlnAlaAspArgArgGln 2032
QY 3534 CCT-----ACCCCAACCCGCTATCTCTATGTCAAAAGGTATGAAA 3572
Db 2033 SerMetAlaPheSerIleLeuAsnThrProLysLysLeuGlyAsnSerLeu---LeuArg 2051
QY 3573 GCTGGAAGCCAGTAGTGGCAGCCTCAGGAGCAGGAGAAATCTGACCAAAATCCAGCCTCGA 3632
Db 2052 ArgGlyAlaSerLysLysAlaLeuSerLysAlaSerProAsnThrArgSerGlyThrArg 2071
QY 3633 GCTGAGACTCAGTCTATGAAAATAGAGCTGAAGAAAATCTGCAGCCAGCAGCAGCTCTCT 3692
Db 2072 -----ArgSerProArgIleAlaThrThrAlaSerAlaAlaThrAlaAlaAla 2088
QY 3693 CTTGGA 3698
Db 2089 IleGly 2090
RESULT 13
US-08-470-950-4
; Sequence 4, Application US/08470950
; Patent No. 5698439
; GENERAL INFORMATION:
; APPLICANT: TOUNKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,950
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-470-950-4
Alignment Scores:
Pred. No.: 7.12e-20 Length: 2101
Score: 409.50 Matches: 293
Percent Similarity: 36.09% Conservative: 249
Best Local Similarity: 19.51% Mismatches: 583
Query Match: 5.30% Indels: 377
DB: 1 Gaps: 59
US-10-788-793-1 (1-4364) x US-08-470-950-4 (1-2101)
QY 51 TTAAGGAGTCGACCAACAGGTGGGAATGAGATCACGAAATCAAGGTGGAGAAAGTTCATCT 110
Db 680 LeuArgSerGluGlnGlnLysAlaThrGluLysGluArgValAlaGlnGluLysAspGln 699

QY 111 ---AACGGGCATGTCTCCTGCCCCCAAGTCCTCCATCATCAGCAGTGTATGGTGGTAAGGCG 167
Db LeuGlnGluGlnLeuGlnAlaLeuLysGluSerLeuLysValThr-----LysGly 716
QY 168 CCTCTGAAGATGCAAAAAGAACAAAGGCCAATCGGAGGAGGAGGATGTCTATGGCTTCC 227
Db SerLeuGluGluGluLysArgArgAlaAlaAspAlaLeuGluGluGlnGlnArgCysIle 736
QY 228 GGAACCTATCAAAAGGCACCTCAAAACCATCTGGAGAAAGTGAGAAAAAGACTAAGAAGTCT 287
Db SerGluLeuLysAlaGluThrArgSerLeuValGluGlnHisLysArgGluArgLysGlu 756
QY 288 GTGGAGTTATCCAAGGAGGAC-----CTCATCCAGCTCCTGACT 326
Db LeuGluGluGluArgAlaGlyArgLysGlyLeuGluAlaArgLeuLeuGlnLeuGlyGlu 776
QY 327 ATCATGGAAGGGAGTTGCAGGCT---CGAGAAGATGTCATCCACATGCTGAGGACAGAG 383
Db AlaHisGlnAlaGluThrGluValLeuArgArgGluLeuAlaGluAlaMetAlaAlaGln 796
QY 384 AAAACCAAG-----CCCGAGGTCTG----- 404
Db HisThrAlaGluSerGluCysGluGlnLeuValLysGluValAlaAlaTrpArgAspGly 816
QY 405 -----GAGGCACACTATGATCTGCAGAACCTCAGAAAAAGTGCTT 443
Db TyrGluAspSerGlnGlnGluAlaGlnTyrGlyAlaMetPheGlnGluGlnLeuMet 836
QY 444 CGGGTC-----CTGCACCGAGATGCCATCCTTGTCTCAAGAGAAGTCC 485
Db ThrLeuLysGluGluCysGluLysAlaArgGlnGlnLeuGlnAlaLysGluLysVal 856
QY 486 ATAGGA----- 491
Db AlaGlyIleGluSerHisSerGluLeuGlnIleSerArgGlnGlnAsnLysLeuAlaGlu 876
QY 492 -----GAAGACGTCTATGAGAAAACCTATC----- 515
Db LeuHisAlaAsnLeuAlaArgAlaLeuGlnGlnValGlnGluLysGluValArgAlaGln 896
QY 516 -----TCAGAGCTGGACAGACTGGAGGAAAAGCAGAAGGAGACGTACCGCCGCA TG 566
Db LysLeuAlaAspAspLeuSerThrLeuGlnGluLysMetAlaAlaThrSerLysGluVal 916
QY 567 -----CTAGAGCAGCTGCTG-----CTGGCTGAGAAGTGTCA CAGGCGC 605
Db AlaArgLeuGluThrLeuValArgLysAlaGlyGluGlnGlnGluThrAlaSerArgGlu 936
QY 606 ACCGTGTACGAGCTGGAGAACGAGAACCAAGCACACTGACTACATG----- 653
Db LeuValLysGluProAlaArgAlaGlyAspArgGlnProGluTrpLeuGluGlnGln 956
QY 654 -----AACAAAGAGCGACGAC 668
Db GlyArgGlnPheCysSerThrGlnAlaAlaLeuGlnAlaMetGluArgGluAlaGluGln 976
QY 669 TTCACCAACCTGCTGGAG----- 686
Db MetGlyAsnGluLeuGluArgLeuArgAlaAlaLeuMetGluSerGlnGlyGlnGlnGln 996
QY 687 -----CAGGACGAGAGAGGTG 704
Db GluGluArgGlyGlnGlnGluArgGluValAlaArgLeuThrGlnGluArgGlyArgAla 1016
QY 705 AAAAAGCTCCTTGAACAGAAAAAGCTTACCAAGCCCGCAAGAAAAAGGAAAAACGCTAAG 764
Db GlnAlaAspLeuAlaLeuGluLysAlaAlaArgAlaGluLeuGluMetArgLeuGlnAsn 1036
QY 765 CGGCTCAACAAACTTCGAGATGAGCTTGTGAAGCTCAAGTCTTCGCCCTCATGTTGGTG 824
Db AlaLeuAsnGluGlnArgValGluPheAlaThrLeuGlnGluAlaLeuAlaHisAlaLeu 1056
QY 825 GACGAGAGGCAGATGCACATCGAGCAACTG-----GGCCTGCAGAGTCAGAAA 872

Db 1057 ThrGluLysGluGlyLysAspGlnGluLeuAlaLysLeuArgGlyLeuGluAlaAlaGln 1076
QY 873 GTCCAGGACCTCACTCAGAAGCTGAGGAGGAGGAAGAAAAACTCAAAGCGGTCACTTAC 932
Db IleLysGluLeu--GluGluLeuArgGlnThrValLysGlnLeuLysGluGlnLeuAla 1095
QY 933 AAATCCAAGGAAGAC----- 947
Db LysLysGluLysGluHisAlaSerGlySerGlyAlaGlnSerGluAlaAlaGlyArgThr 1115
QY 948 -----CGCCAGAAGCTGCTCAAGTTAGAAAGTGGAC 977
Db GluProThrGlyProLysLeuGluAlaLeuArgAlaGluValSerLysLeuGluGlnGln 1135
QY 978 TTCGAACACAAAGGCCTCGAGGTTTTCCAGGAGGACACGAAGAGATGAACGCCAAAATTGGCG 1037
Db CysGlnLysGlnGlnGluGlnAlaAspSerLeuGluArgSerLeuGluAlaGluArgAla 1155
QY 1038 AATCAAGAATCTCACAAACGGCAACTTCGACTCAAACTGGTTGGCTTATCGCAAAGGATT 1097
Db SerArgAlaGluArgAspSerAlaLeuGluThrLeuGlnGlyGlnLeuGluLysAla 1175
QY 1098 GAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAGCGCAGAGGAAGAGCTCCAGGAGCTG 1157
Db GlnGluLeuGlyHisSerGlnSerAlaLeuAlaSerAlaGlnArgGluLeuAlaAlaPhe 1195
QY 1158 AGAGAGAAA-----ATTGCCAAA 1175
Db ArgThrLysValGlnAspHisSerLysAlaGluAspGluTrpLysAlaGlnValAlaArg 1215
QY 1176 GGG-----GAATGTGGAAC-----TCAGTCTCATGGCGGAAGTGGAG 1214
Db GlyArgGlnGluAlaGluArgLysAsnSerLeuIleSerSerLeuGluGluValSer 1235
QY 1215 AGTCTGCGCAAGCGCGTGTCTTGAGATGGAGGGCAAGGATGAAGAGATC----- 1262
Db IleLeuAsnArgGlnValLeuGluLysGluGlyGluSerLysGluLeuLysArgLeuVal 1255
QY 1263 -----ACGAAGACCGAGGCCCGCTGC----- 1283
Db MetAlaGluSerGluLysSerGlnLysLeuGluGluSerCysAlaCysCysArgGlnArg 1275
QY 1283 ----- 1283
Db GlnProAlaThrValProGluLeuGlnAsnAlaAlaLeuLeuCysGlyArgArgCysArg 1295
QY 1284 -----CGGGAGCTGAAGAAGAAAGCTCCAAGAGGAGGAACAACACCACAGCAAGGAAC TT 1334
Db AlaSerGlyArgGluAlaGluLysGlnArgValAlaSerGluAsnLeuArgGlnGluLeu 1315
QY 1335 AGACTAGAAGTGGAGAGCTGCAGAAGAGGATGTCTGAGTGGAGAAAGCTGAGGAAGCG 1394
Db ThrSerGlnAlaGluArgAlaGluGluLeuGlyGlnGluLeuLysAlaTrpGlnGluLys 1335
QY 1395 TTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAGGAGAAGAAC 1454
Db PhePheGlnLysGluGlnAlaLeuSerThrLeuGlnLeu-----GluHisThr 1351
QY 1455 CTAAACAAAAGACCTGCTCAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAA 1514
Db SerThrGlnAlaLeuValSerGluLeuLeuProAlaLysHisLeuCysGlnGlnLeuGln 1371
QY 1515 TGCTCCGAGAGTAGACTGGAGAAAGGCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTG 1574
Db AlaGluGlnAlaAlaGluLysArgHis----- 1381
QY 1575 AAGTCCTTCACTGTGTGCTGGTGGATGAGAGGAAAAATATGATGGAGAAAAATAAGCAA 1634
Db -----ArgGluGluLeuGluGln 1387
QY 1635 GAAGAGAGGAAAAGTGGATGGTTGAATAAAAAACTTT-----AAGTGGAGCAGGGA 1685

Db 1388 SerLysGlnAlaAlaGlyGlyLeuArgAlaGluLeuLeuArgAlaGlnArgGluLeuGly 1407
QY 1686 AAAGTCATGGATGTGACGGAAGAAAGCTAATCGAGGAAGCAAGAAAGCTTTTAAAACTCAAA 1745
Db 1408 GluLeuIleProLeuArgGlnLysValAlaGluGlnArgThrAlaGlnGlnLeuArg 1427
QY 1746 TCTGAAATGGAGGAA-----AAGGAGTACAGTCTGACAAAAGGAGAGGGATGAGCTGATG 1799
Db 1428 AlaGluLysAlaSerTyrAlaGluGlnLeuSerMetLeuLysLysAlaHisGlyLeuLeu 1447
QY 1800 GGTAAACTGAGGAGCGAAGAAGAAAGGTCTGTGAACTGAGCTGCAGTGTAGACTTACTA 1859
Db 1448 -----AlaGluGluAsnArgGly-----Leu 1454
QY 1860 AAGAAGCGGCTTGAT---GGCATAGAGGAGGTAGAAAAGGGAATAAAACCGAGGTAGGTG 1916
Db 1455 GlyGluArgAlaAsnLeuGlyArgGlnPheLeuGluValGluLeuAspGlnAlaArgGlu 1474
QY 1917 TGCAAGGGTCTGAGTTCACC-----TGCCCGGAAGACAATAAGATCAGAGAACTA 1967
Db 1475 LysTyrValGlnGluLeuAlaAlaValArgAlaAspAlaGluThrArgLeuAlaGluVal 1494
QY 1968 AGCTTTGAAATCGAGAGACTGAAGAAACGGCTCCAGCAGATTGGAGGTG---GTGGAGGGG 2024
Db 1495 GlnArgGluAlaGlnSerThrAlaArgGluLeuGluValMetThrAlaLysTyrGluGly 1514
QY 2025 GACTTGTATGAAGACCGAGGACGAATATGACCAGTTGGAGCAGAAGTTTCAGAACCGAGCAG 2084
Db 1515 AlaLysValLysValLeuGluGlu-----ArgGlnArgPheGlnGluGluArg 1530
QY 2085 GATAAGGCAAACTTCCTCTCCAGCAGCTCGAGGAATCAAAACACCAAAATGGSCAAGCAC 2144
Db 1531 GlnLys-----LeuThrAlaGlnValGluGluLeuSerLysLysLeuAlaAspSer 1547
QY 2145 AAAGCCATAGAGAAAGGGAGGCGGTGAGCCAGCAAGCCGAACCTGCGACACAGGTTTCGG 2204
Db 1548 -----AspGlnAlaSerLysValGlnGlnGln-----LysLeuLys 1559
QY 2205 CTGGAGGAGGCTAAAGTCGTGATTTACAGGCGCAGGTGCAGGCTCTCAAGGAGAAGATC 2264
Db 1560 AlaValGlnAlaGlnGlyGlyGluSerGlnGlnGluAlaGlnArgPheGlnAlaGlnLeu 1579
QY 2265 CACGAGCTGATGAACAAGGAAGACCAAGCTGTCTCAGCTCCAAGTCGACTATTCGGTCTCT 2324
Db 1580 AsnGluLeuGlnAlaGlnLeuSerGlnLysGluGlnAlaAlaGluHisTyrLysLeuGln 1599
QY 2325 CAGCAAAAGATTATGGAAGAAGAAACTAAGAACAAAGAACATGGGGAGGGAGTCTCTCAAT 2384
Db 1600 MetGluLysAlaLysThrHisTyrAspAlaLysLysGlnGlnAsnGlnGluLeuGlnGlu 1619
QY 2385 CTGACCAAGGAGCTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAAC 2444
Db 1620 GlnLeuArgSerLeuGluGlnLeuGlnLysGluAsnLysGluLeuArgAlaGluAla--- 1638
QY 2445 GGCCGAAGGATGTGGACGTGCCTGTGGCCCTCCACTGGGTGCAGACC-----GAG 2495
Db 1639 -----GluArgLeuGlyHisGluLeuGlnGlnAlaGlyLeuLysThrLysGluAlaGlu 1656
QY 2496 GCGGTGTGCGGGATGCTGCGGAGGAGAGACCCCGGCTGTGTTTCATTCGAAAATCCTTC 2555
Db 1657 GlnThrCysArgHisLeuThrAlaGln-----ValArgSerLeuGlu 1670
QY 2556 CAGGAGGAAATCACATCATGATGATTAATCTTCGACAGGTAGGCCTGAAGAAACCCATGAA 2615
Db 1671 AlaGlnValAlaHisAlaAspGlnGlnLeuArgAspLeuGly---LysPheGlnValAla 1689
QY 2616 CGGTCTCGGTCTCGACAGGTATCCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCT 2675
Db 1690 ThrAspAlaLeuLysSerArgGluProGlnAlaLysProGlnLeuAspLeuSerIleAsp 1709
QY 2676 TGGATTCTTGGATGAGAAAAAGAGAAAAACGGTCTCTCCACTCCGACGAGAGAAAGGCC 2735
Db 1710 SerLeuAsp---LeuSerCysGluGluGlyThrProLeuSerIleThrSerLysLeuPro 1728

QY 2736 AGGCCAAAC---CAGGGTGCAGGGCACCCCGGGAGCTGGTC----- 2774
Db 1729 ArgThrGlnProAspGlyThrSerValProGlyGluProAlaSerProIleSerGlnArg 1748
QY 2775 CTAGCACCAAAAGCAGGGCCAGCCCTACACATCCGTGTGACACCAGATCATGAGAACAGC 2834
Db 1749 LeuProProLysValGluSerLeuGluSerLeuTyrPheThrProIleProAlaArgSer 1768
QY 2835 ACTGCCACCCTGGAGATCACAAGCCCCCACATCTGAAGAGTTTTTC----- 2879
Db 1769 GlnAlaProLeuGluSerSerLeuAspSerLeuGlyAspValPheLeuAspSerGlyArg 1788
QY 2880 -----TCTAGTACCACCGTCATTCTCTACCTTAGGCAACAGAAA 2918
Db 1789 LysThrArgSerAlaArgArgThrThrGlnIleIleAsnIleThrMetThrLysLys 1808
QY 2919 CCAAGAATAACCATATTCCATCACCCCAATGTCTGTGCGAAAAGCCCCAAAGTGCAGAT 2978
Db 1809 LeuAspValGluProAspSerAlaAsnSerSerPheTyrSerThrArgSerAlaPro 1828
QY 2979 CCTACTCTCGGCCCAGAACGAGCCATGTCCCTGTGCACGATTACTACTATTTCCAGAGAG 3038
Db 1829 AlaSerGlnAlaSerLeuArgAlaThrSer-----SerThrGlnSerLeuAlaArgLeu 1846
QY 3039 AAGAGCCCCGGAAGGTGGAAGGAGCGCCTTTGCCGACAGGCCTGCA---TCCCCCATCCAA 3095
Db 1847 GlySerProAspTyrGlyAsnSerAlaLeuLeuSerLeuProGlyTyrArgPro----- 1864
QY 3096 ATCATGACGGTGTCAACATCTGCAGCTCCCACTGAAATCGCTGTCTCTCTGAAATCTCAG 3155
Db 1865 -----ThrThrArgSerSerAlaArgArgSerGlnAlaGlyVal-----SerSer 1879
QY 3156 GAAGTGCCTATGGGAAGGACTATCTCTCAAAGTCAACCCGGAAAAACAACACTGTTCAGCC 3215
Db 1880 GlyAlaProProGlyArgAsnSerPheTyrMet--GlyThrCysGlnAspGluProGlu 1898
QY 3216 CCGGTGCGGAAGTACAACTCCAATGCTAATATATCATCACCCGGAAGACATAAAATT--- 3272
Db 1899 GlnLeuAspAspTrpAsnArgIleAlaGluLeu-----GlnGlnArgAsnArgValCys 1916
QY 3273 -----CACATTCACTGGGTTCTCAGTTTAAGCGATCTCTGGGGCTGCCGCTGAAGGC 3326
Db 1917 ProProHisLeuLysThrCysTyrProLeuGluSerArgProSerLeuSerLeuGlyThr 1936
QY 3327 GTGAGC-----CCAGTTATCACCGTCCCG----- 3350
Db 1937 IleThrAspGluGluMetLysThrGlyAspProGlnGluThrLeuArgAlaSerMet 1956
QY 3351 ---CCTGTCAAGTGCAGCGGAGAGGAGTTTCTACAGGCACAGTCTCTCGTCTCTCCC 3407
Db 1957 GlnProIleGlnIle---AlaGlu-----GlyThrGlyIleThrThrArgGlnGln 1972
QY 3408 AGGAACCACTCTCT-----TCAAGACCCCGTGTCTAGCAAAAGTG 3446
Db 1973 ArgLysArgValSerLeuGluProHisGlnGlyProGlyThrProGluSerLysAla 1992
QY 3447 ACCAGCACTATA-----ACTATAACCCCG-----GTC 3473
Db 1993 ThrSerCysPheProArgProMetThrProArgAspArgHisGluGlyArgLysGlnSer 2012
QY 3474 ACAACGTCTATCCACACGAGGAACCCCAATCAGTGTGAGGACAAGATGGGTCTATCTCAGCGG 3533
Db 2013 ThrThrGluAlaGlnLysLysAlaAlaProAlaSerThrLysGlnAlaAspArgGln 2032
QY 3534 CCT-----ACCCCAACCCCGCATTCCTATGTCAAAAGGTATGAAA 3572
Db 2033 SerMetAlaPheSerIleLeuAsnThrProLysLysLeuGlyAsnSerLeu---LeuArg 2051
QY 3573 GCTGGAAGCCAGTAGTGGCAGCCTCAGGAGCAGGAAATCTGACCAAAATTCAGCCTCGA 3632
Db 2052 ArgGlyAlaSerLysLysAlaLeuSerLysAlaSerProAsnThrArgSerGlyThrArg 2071

QY 948 -----CGCCAGAAGCTGCTCAAGTTAGAAGTGGAC 977
Db GluProThrGlyProLysLeuGluAlaLeuArgAlaGluValSerLysLeuGluGln 1135
QY 978 TTCGAACACAAAGCCTCGAGGTTTCCAGGAGCAGCAAGAGATGAACGCCAAATTTGGCG 1037
Db CysGlnLysGlnGluGlnAlaAspSerLeuGluArgSerLeuGluAlaGluArgAla 1155
QY 1038 AATCAAGAAATCTCAACACCGGCAACTTCGACTCAAACTGGTTGGCTTATCGCAAAGGATT 1097
Db SerArgAlaGluArgAspSerAlaLeuGluThrLeuGlnGlyGlnLeuGluLysAla 1175
QY 1098 GAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAAGCAGAGAGAGTCCAGGAGCTG 1157
Db GlnGluLeuGlyHisSerGlnSerAlaLeuAlaSerAlaGlnArgGluLeuAlaAlaPhe 1195
QY 1158 AGAGAGAAA-----ATTGCCAAA 1175
Db ArgThrLysValGlnAspHisSerLysAlaGluAspGluTrpLysAlaGlnValAlaArg 1215
QY 1176 GGG-----GAATGTGGAAC-----TCCAGTCTCATGGCGGAAGTGGAG 1214
Db GlyArgGlnGluAlaGluArgLysAsnSerLeuIleSerSerLeuGluGluValSer 1235
QY 1215 AGTCTGCGCAAGCGGTGCTTGAGATGGAGGGCAGGATGAAGATC----- 1262
Db IleLeuAsnArgGlnValLeuGluLysGluGlyGluSerLysGluLeuLysArgLeuVal 1255
QY 1263 -----ACGAAGACCGAGGCCAGTGC----- 1283
Db MetAlaGluSerGluLysSerGlnLysLeuGluGluSerCysAlaCysCysArgGlnArg 1275
QY 1283 ----- 1283
Db GlnProAlaThrValProGluLeuGlnAsnAlaAlaLeuLeuCysGlyArgArgCysArg 1295
QY 1284 -----CGGGAGCTGAAGAAGAAAGCTCCAGAGGAAGAACACACAGCAAGAACTT 1334
Db AlaSerGlyArgGluAlaGluLysGlnArgValAlaSerGluAsnLeuArgGlnGluLeu 1315
QY 1335 AGACTAGAAGTGGAGAAGCTGCAGAAGAGGATGTCTGAGCTGGAGAAAGCTGGAGGAAGCG 1394
Db ThrSerGlnAlaGluArgAlaGluGluLeuGlyGlnGluLeuLysAlaTrpGlnGluLys 1335
QY 1395 TTCAGCCGGAGTAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAGGAAGAAC 1454
Db PhePheGlnLysGluGlnAlaLeuSerThrLeuGlnLeu-----GluHisThr 1351
QY 1455 CTAACCAAAGACCTGCTGAACGAGCTGGAGGTGGTCAAGAGTCGAGTTAAAGAACTCGAA 1514
Db SerThrGlnAlaLeuValSerGluLeuLeuProAlaLysHisLeuCysGlnGlnLeuGln 1371
QY 1515 TGCTCCGAGAGTAGCTGGAGAAGGCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTG 1574
Db AlaGluGlnAlaAlaAlaGluLysArgHis----- 1381
QY 1575 AAGTCCTTCACTGTGATGCTGGTGGATGAGAGGAAAATATGATGGAGAAAATAAGCAA 1634
Db -----ArgGluGluLeuGluGln 1387
QY 1635 GAAGAGAGGAAAGTGGATGGGTTGAATAAAAACTTT-----AAGGTGGAGCAGGGA 1685
Db SerLysGlnAlaAlaGlyLeuArgAlaGluLeuLeuArgAlaGlnArgGluLeuGly 1407
QY 1686 AAAGTCATGGATGTACGGAAAAAGCTAATCGAGGAAAAGCAAGACTTTTAAAACTCAAA 1745
Db GluLeuIleProLeuArgGlnLysValAlaGluGlnGluArgThrAlaGlnGlnLeuArg 1427
QY 1746 TCTGAAATGGAGGAA-----AAGGAGTACAGTCTGACAAAAGGAGGGATGAGCTGATG 1799
Db AlaGluLysAlaSerTyrAlaGluGlnLeuSerMetLeuLysLysAlaHisGlyLeuLeu 1447

QY 1800 GGTAAACTGAGGAGCGAAGAAAGGTCTGTGAACTGAGCTGCAGCTGTAGACTTACTA 1859
Db -----AlaGluGluAsnArgGly-----Leu 1454
QY 1860 AAGAAGCGGCTTGAT--GGCATAGAGGAGGTAGAAAGGAAATAAACCCGAGGTAGGTGCG 1916
Db GlyGluArgAlaAsnLeuGlyArgGlnPheLeuGluValGluLeuAspGlnAlaArgGlu 1474
QY 1917 TGCAAGGGGTCTGAGTTCACC-----TGCCCCGAAGACAAATAAGATCAGAGAACTA 1967
Db LysTyrValGlnGluLeuAlaAlaValArgAlaAspAlaGluThrArgLeuAlaGluVal 1494
QY 1968 ACGCTTGAATCGAGAGACTGAAGAAACCGCTCCAGCAGTTGGAGGTG---GTGGAGGGG 2024
Db GlnArgGluAlaGlnSerThrAlaArgGluLeuGluValMetThrAlaLysTyrGluGly 1514
QY 2025 GACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAGCAGAAAGTTCAACACCGAGCAG 2084
Db AlaLysValLysValLeuGluGlu-----ArgGlnArgPheGlnGluGluArg 1530
QY 2085 GATAAGGCAAACTCTCTCTCCAGCAGCTCGAGGAAATCAAAACACCAATGGCCAAAGCAC 2144
Db GlnLys-----LeuThrAlaGlnValGluGluLeuSerLysLysLeuAlaAspSer 1547
QY 2145 AAAGCCATAGAGAAAGGGAGGCGCTGAGCGAGGAAAGCCGAACCTGCGACACAGGTTTCGG 2204
Db -----AspGlnAlaSerLysValGlnGlnGln-----LysLeuLys 1559
QY 2205 CTGGAGGAGGCTAAAAAGTCGTGATTTACAGCGCGAGGTGCAGGCTCTCAAGGAGAAAGATC 2264
Db AlaValGlnAlaGlnGlyGlyGluSerGlnGlnGluAlaGlnArgPheGlnAlaGlnLeu 1579
QY 2265 CACGAGCTGATGAACAAGGAAGACCAAGCTGTCTCAGCTCCAAGTCGACTATTCGGTCCCTT 2324
Db AsnGluLeuGlnAlaGlnLeuSerGlnLysGluGlnAlaAlaGluHisTyrLysLeuGln 1599
QY 2325 CAGCAAAGATTATGGAAGAAGAAACTAAAGAACAAACATGGGAGGAGGCTCCTCAAT 2384
Db MetGluLysAlaLysThrHisTyrAspAlaLysLysGlnGlnAsnGlnGluLeuGlnGlu 1619
QY 2385 CTGACCAAGGAGCTAGAGCTTTCCAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAAC 2444
Db GlnLeuArgSerLeuGluGlnLeuGlnLysGluAsnLysGluLeuArgAlaGluAla--- 1638
QY 2445 GGCCGAAGGATGGTGGACGTGCCTGTGGCCTCCACTGGGTGCAGACC-----GAG 2495
Db -----GluArgLeuGlyHisGluLeuGlnGlnAlaGlyLeuLysThrLysGluAlaGlu 1656
QY 2496 GCGGTGTGCGGGGATGCTGCGGAGGAGGAGACCCCGGCTGTGTTTCATTCGCAATCCTTC 2555
Db GlnThrCysArgHisLeuThrAlaGln-----ValArgSerLeuGlu 1670
QY 2556 CAGGAGGAAAATCAGATCATGATGTAATCTTCAGCAGGTAGGCCTGAAGAAACCCATGGAA 2615
Db AlaGlnValAlaHisAlaAspGlnGlnLeuArgAspLeuGly---LysPheGlnValAla 1689
QY 2616 CGGTCTCGTCTCGACAGGTATCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCT 2675
Db ThrAspAlaLeuLysSerArgGluProGlnAlaLysProGlnLeuAspLeuSerIleAsp 1709
QY 2676 TGGATTCTCTGGATGAGAAAAAGAAAACGGTCTCTCCACTCCGAGGAGAAAGGGCCC 2735
Db SerLeuAsp---LeuSerCysGluGluGlyThrProLeuSerIleThrSerLysLeuPro 1728
QY 2736 AGGCCAAAC---CAGGGTGCAGGGCACCCCGGGGAGCTGGTC----- 2774
Db ArgThrGlnProAspGlyThrSerValProGlyGluProAlaSerProIleSerGlnArg 1748
QY 2775 CTAGCACCAAGCAGGGGCCAGCCCCCTACACATCCGTGTGACACCAGATCATGAGAACAGC 2834
Db LeuProProLysValGluSerLeuGluSerLeuTyrPheThrProIleProAlaArgSer 1768
QY 2835 ACTGCCACCCCTGGAGATCACAAGCCCCCACATCTGAAGAGTTTTTC----- 2879

Db 1769 GlnAlaProLeuGluSerSerLeuAspSerLeuGlyAspValPheLeuAspSerGlyArg 1788
QY 2880 -----TCTAGTACCACCGTCACTTCTTACCTTAGGCAACCCAGAAA 2918
Db 1789 LysThrArgSerAlaArgArgThrThrGlnIleAsnIleThrMetThrLysLys 1808
QY 2919 CCAAGAATAACCATTTCCCATCACCCAAATGTCTGCGAAAAAGCCCCAAAGTGCAGAT 2978
Db 1809 LeuAspValGluGluProAspSerAlaAsnSerSerPheTyrSerThrArgSerAlaPro 1828
QY 2979 CCTACTCTCGGCCCCAGAACGAGCCATGTCCCTGTCTCACGATTACTACTATTTCAGAGAG 3038
Db 1829 AlaSerGlnAlaSerLeuArgAlaThrSer-----SerThrGlnSerLeuAlaArgLeu 1846
QY 3039 AAGAGCCCGGAAGGTGGAAGGAGCGCCTTTGCCGACAGGCTGCA---TCCCCCATCCAA 3095
Db 1847 GlySerProAspTyrGlyAsnSerAlaLeuLeuSerLeuProGlyTyrArgPro----- 1864
QY 3096 ATCATGACGGTGTCAACATCTGCAGCTCCCACTGAAATCGCTGTCTCTCTGAATCTCAG 3155
Db 1865 -----ThrThrArgSerSerAlaArgSerGlnAlaGlyVal-----SerSer 1879
QY 3156 GAAGTGCTATGGGAAGGACTATCCTCAAAGTCAACCCCGAAAAACAACAACTGTTCCAGCC 3215
Db 1880 GlyAlaProProGlyArgAsnSerPheTyrMet---GlyThrCysGlnAspGluProGlu 1898
QY 3216 CCGGTGCGGAAGTACAACTCCAATGCTAATATCATCACCGGAAGACATAAAATT--- 3272
Db 1899 GlnLeuAspAspTrpAsnArgIleAlaGluLeu-----GlnGlnArgAsnArgValCys 1916
QY 3273 -----CACATTCACTGGGTCTCAGTTAAGCGATCTCTGGGCGCTGCCGCTGAAGGC 3326
Db 1917 ProProHisLeuLysThrCysTyrProLeuGluSerArgProSerLeuSerLeuGlyThr 1936
QY 3327 GTGAGC-----CCAGTTATCACCGTCCCG----- 3350
Db 1937 IleThrAspGluGluMetLysThrGlyAspProGlnGluThrLeuArgAlaSerMet 1956
QY 3351 ---CCTGTCAACGTGACAGCGGAGAGGTTTCTACAGGCACAGTCTCTCGTCTCCC 3407
Db 1957 GlnProIleGlnIle---AlaGlu-----GlyThrGlyIleThrThrArgGlnGln 1972
QY 3408 AGGAACCACTCTCT-----TCAGACCCCGGTGCTAGCAAAAGTG 3446
Db 1973 ArgLysArgValSerLeuGluProHisGlnGlyProGlyThrProGluSerLysLysAla 1992
QY 3447 ACCAGCACTATA-----ACTATAACCCCG-----GTC 3473
Db 1993 ThrSerCysPheProArgProMetThrProArgAspArgHisGluGlyArgLysGlnSer 2012
QY 3474 ACAACGTCAATCCACACGAGGAACCCAAATCAGTGTACGACACAAGATGGGTCACTCAGCGG 3533
Db 2013 ThrThrGluAlaGlnLysLysAlaAlaProAlaSerThrLysGlnAlaAspArgArgGln 2032
QY 3534 CCT-----ACCCCAACCCGCATTCCTCTATGTCAAAAGGTATGAAA 3572
Db 2033 SerMetAlaPheSerIleLeuAsnThrProLysLysLeuGlyAsnSerLeu---LeuArg 2051
QY 3573 GCTGGAAGCCAGTAGTGGCAGCCTCAGGAGCAGGAAATCTGACCAAAATTCAGCCCTCGA 3632
Db 2052 ArgGlyAlaSerLysLysAlaLeuSerLysAlaSerProAsnThrArgSerGlyThrArg 2071
QY 3633 GCTGAGACTCAGTCTATGAAAATAGAGCTGAAGAAATCTGCAGCCAGCAGCACTGCCTCT 3692
Db 2072 -----ArgSerProArgIleAlaThrThrThrAlaSerAlaAlaThrAlaAlaAla 2088
QY 3693 CTTGGA 3698
Db 2089 IleGly 2090

RESULT 15
US-08-195-487-4

; Sequence 4, Application US/08195487
; Patent No. 5783403
; GENERAL INFORMATION:
; APPLICANT: TOURKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,487
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/901,701
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ. EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-195-487-4
Alignment Scores:
Pred. No.: 7.12e-20 Length: 2101
Score: 409.50 Matches: 293
Percent Similarity: 36.09% Conservative: 249
Best Local Similarity: 19.51% Mismatches: 583
Query Match: 5.30% Indels: 377
DB: 1 Gaps: 59
US-10-788-793-1 (1-4364) x US-08-195-487-4 (1-2101)

QY 51 TTAAGGAGTCGACAACAGGTGGGAATGAGATCACGAAATCAAGGTGGAGAAAAGTTCATCT 110
Db 680 LeuArgSerGluGlnGlnLysAlaThrGluLysGluArgValAlaGlnGluLysAspGln 699
QY 111 ---AACGGGCATGTCTCTGCTCCCAAGTCTCCATCATCAGCAGTGTGTTGTAAGGGC 167
Db 700 LeuGlnGluGlnLeuGlnAlaLeuLysGluSerLeuLysValThr-----LysGly 716
QY 168 CCCTCAGAAGATGCAAAAAAGAACAAAGCCCAATCGGAAGGAGGAGATGTCTGCTTCC 227
Db 717 SerLeuGluGluGluLysArgArgAlaAlaAspAlaLeuGluGluGlnGlnArgCysIle 736
QY 228 GGAACATCAAAAGGCGACCTCAAACCATCTGGAGAAAGTGAGAAAAAGACTAAGAAAGTCT 287
Db 737 SerGluLeuLysAlaGluThrArgSerLeuValGluGlnHisLysArgGluArgLysGlu 756
QY 288 GTGGAGTTATCCAAGGAGGAC-----CTCATCCAGCTCCTGAGT 326
Db 757 LeuGluGluGluArgAlaGlyArgLysGlyLeuGluAlaArgLeuLeuGlnLeuGlyGlu 776
QY 327 ATCATGGAAGGGGAGTTGTCAGGCT---CGAGAAGAATGTCTATCCCATCTGCTGAGGACAGAG 383

Db 777 AlaHisGlnAlaGluThrGluValLeuArgArgGluLeuAlaGluAlaMetAlaAlaGln 796
QY 384 AAAACCAAG-----CCCGAGGTTCTG----- 404
Db 797 HisThrAlaGluSerGluCysGluGlnLeuValLysGluValAlaAlaTrpArgAspGly 816
QY 405 -----GAGGCACACTATGATCTGCAGAACCTGAGAAAAGTGTCTT 443
Db 817 TyrGluAspSerGlnGlnGluAlaGlnTyrGlyAlaMetPheGlnGluGlnLeuMet 836
QY 444 CGGGTC-----CTGCACCGAGATGCCATCCTTGCTCAAGAGAAGTCC 485
Db 837 ThrLeuLysGluGluCysGluLysAlaArgGlnGluLeuGlnGluAlaLysGluLysVal 856
QY 486 ATAGGA----- 491
Db 857 AlaGlyIleGluSerHisSerGluLeuGlnIleSerArgGlnGlnAsnLysLeuAlaGlu 876
QY 492 -----GAAGACGTCCTATGAGAAACCTATC----- 515
Db 877 LeuHisAlaAsnLeuAlaArgAlaLeuGlnGlnValGlnLysGluValArgAlaGln 896
QY 516 -----TCAGAGCTGGACAGACTGGAGGAAAAGCAGAAAGGAGACGTACCGCCGATG 566
Db 897 LysLeuAlaAspAspLeuSerThrLeuGlnGlnLysMetAlaAlaThrSerLysGluVal 916
QY 567 -----CTAGAGCAGCTGCTG-----CTGGCTGAGAAAGTGTACAGGGCG 605
Db 917 AlaArgLeuGluThrLeuValArgLysAlaGlyGluGlnGlnGluThrAlaSerArgGlu 936
QY 606 ACCGTGTACGAGCTGGAGAACGAGAAAGCAGACACACTGACTACATG----- 653
Db 937 LeuValLysGluProAlaArgAlaGlyAspArgGlnProGluTrpLeuGluGlnGln 956
QY 654 -----AACAGAGCGCAGCAG 668
Db 957 GlyArgGlnPheCysSerThrGlnAlaAlaLeuGlnAlaMetGluArgGluAlaGluGln 976
QY 669 TTCACCAACCTGCTGGAG-----AACAGAGCGCAGCAG 686
Db 977 MetGlyAsnGluLeuGluArgLeuAlaAlaLeuMetGluSerGlnGlyGlnGlnGln 996
QY 687 -----CAGGAGCGAGAGAGGTTG 704
Db 997 GluGluArgGlyGlnGlnGluArgGluValAlaArgLeuThrGlnGluArgGlyArgAla 1016
QY 705 AAAAAGCTCCTTGAACAAAGAAAAGCTTACCAGCCCGCAAAGAAAAGGAAAACGCTAAG 764
Db 1017 GlnAlaAspLeuAlaLeuGluLysAlaAlaArgAlaGluLeuGluMetArgLeuGlnAsn 1036
QY 765 CGGCTCAACAACTCTCGAGATGAGCTTGTGAAGCTCAAGTCCTTCGCCCTCATGTTGGTG 824
Db 1037 AlaLeuAsnGluGlnArgValGluPheAlaThrLeuGlnGluAlaLeuAlaHisAlaLeu 1056
QY 825 GACGAGAGGCGAGATGCACATCGAGCAACTG-----GGCCTGCAGAGTCAGAAA 872
Db 1057 ThrGluLysGluGlyLysAspGlnGluLeuAlaLysLeuArgGlyLeuGluAlaAlaGln 1076
QY 873 GTCCAGGACCTCACTCAGAAAGCTGAGGGAGGAGGAGAAAACCTCAAGCGGTCACTTAC 932
Db 1077 IleLysGluLeu---GluGluLeuArgGlnThrValLysGlnLeuLysGluGlnLeuAla 1095
QY 933 AAATCCAAGGAAGAC----- 947
Db 1096 LysLysGluLysGluHisAlaSerGlySerGlyAlaGlnSerGluAlaAlaGlyArgThr 1115
QY 948 -----CGCCAGAAGCTGCTCAAGTTAGAAAGTGGAC 977
Db 1116 GluProThrGlyProLysLeuGluAlaLeuArgAlaGluValSerLysLeuGluGlnGln 1135
QY 978 TTCGAACACAAAGCCTCGAGGTTTCCCGAGGACGAGAGATGAACGCCAAATTGGCG 1037

1136 CysGlnLysGlnGlnGluAlaAspSerLeuGluArgSerLeuGluAlaGluArgAla 1155
QY 1038 AATCAAGAATCTCACAAACCGCAACTTCGACTCAAACTGGTTGGCTTATCGCAAAGGATT 1097
Db 1156 SerArgAlaGluArgAspSerAlaLeuGluThrLeuGlnGlyGlnLeuGluGluLysAla 1175
QY 1098 GAGGAGCTGGAAGAGACCAATAAAAGCCTTCAGAAGGCGAGAGAGCTCCAGGAGCTG 1157
Db 1176 GlnGluLeuGlyHisSerGlnSerAlaLeuAlaSerAlaGlnArgGluLeuAlaAlaPhe 1195
QY 1158 AGAGAGAAA-----ATTGCCAAA 1175
Db 1196 ArgThrLysValGlnAspHisSerLysAlaGluAspGluTrpLysAlaGlnValAlaArg 1215
QY 1176 GGG-----GAATGTGGAAC-----TCCAGTCTCATGGCGAAGTGGAG 1214
Db 1216 GlyArgGlnGluAlaGluArgLysAsnSerLeuIleSerSerLeuGluGluValSer 1235
QY 1215 AGTCTGCGCAAGCGCTGCTTGAGATGGAGCGCAAGGATGAAGAGATC----- 1262
Db 1236 IleLeuAsnArgGlnValLeuGluLysGluGlyGluSerLysGluLeuLysArgLeuVal 1255
QY 1263 -----ACGAAGACCGAGGCCAGTGC----- 1283
Db 1256 MetAlaGluSerGluLysSerGlnLysLeuGluSerCysAlaCysCysArgGlnArg 1275
QY 1283 ----- 1283
Db 1276 GlnProAlaThrValProGluLeuGlnAsnAlaAlaLeuLeuCysGlyArgCysArg 1295
QY 1284 -----CGGAGCTGAAGAAGAAGCTCCAAGAGGAAGAACACCACAGCAAGGAACCTT 1334
Db 1296 AlaSerGlyArgGluAlaGluLysGlnArgValAlaSerGluAsnLeuArgGlnGluLeu 1315
QY 1335 AGACTAGAAGTGGAGAAAGCTGCAGAAAGAGGATGTCTGAGCTGGAGAAAGCTGGAAGCG 1394
Db 1316 ThrSerGlnAlaGluArgAlaGluGluLeuGlyGlnGluLysAlaTrpGlnGluLys 1335
QY 1395 TTCAGCCGGAGTAAAGTCGGAATGCACCCAGCTCCATCTGAACCTGGAGAAGGAGAAGAAC 1454
Db 1336 PhePheGlnLysGluGlnAlaLeuSerThrLeuGlnLeu-----GluHisThr 1351
QY 1455 CTAACCAAAAGACCTGCTGAACGAGCTGGAGTGGGTCAAGAGTCGAGTTAAAGAACTCGAA 1514
Db 1352 SerThrGlnAlaLeuValSerGluLeuLeuProAlaLysHisLeuCysGlnGlnLeuGln 1371
QY 1515 TGCTCCGAGAGTAGACTGGAGAAAGGCCGAGTTAAGCCTCAAAGATGACCTTACAAAGCTG 1574
Db 1372 AlaGluGlnAlaAlaAlaGluLysArgHis----- 1381
QY 1575 AAGTCCTTCACTGTGTGTGCTGGTGTGATGAGAGGAAAAATATGATGGAGAAAAATAAGCAA 1634
Db 1382 -----ArgGluGluLeuGluGln 1387
QY 1635 GAAGAGAGGAAAGTGGATGGTGTGAATAAAAACTTT-----AAGGTGGAGCAGGGA 1685
Db 1388 SerLysGlnAlaAlaGlyGlyLeuArgAlaGluLeuLeuArgAlaGlnArgGluLeuGly 1407
QY 1686 AAAGTCATGGATGTGACGGAAGTAATTCGAGGAAAGCAAGAGCTTTTAAAACTCAAA 1745
Db 1408 GluLeuIleProLeuArgGlnLysValAlaGluGlnAspArgThrAlaGlnGlnLeuArg 1427
QY 1746 TCTGAAATGGAGGAA-----AAGGAGTACAGTCTGACAAAGGAGAGGATGAGCTGATG 1799
Db 1428 AlaGluLysAlaSerTyrAlaGluGlnLeuSerMetLeuLysLysAlaHisGlyLeuLeu 1447
QY 1800 GGTAACCTGAGGAGCGAAGAAAGGTCTCTGTAAGTGTGAGCTGACAGTGTAGACTTACTA 1859
Db 1448 -----AlaGluGluAsnArgGly-----Leu 1454
QY 1860 AAGAAGCGGCTTGAT---GGCATAGAGAGGCTAGAAAGGGAATAAACCGAGGTAGGTGCG 1916
Db 1455 GlyGluArgAlaAsnLeuGlyArgGlnPheLeuGluValGluLeuAspGlnAlaArgGlu 1474

QY	1917	TGCAAGGGGTCTGAGTTCACC-----TGCCCGGAAGACAATAAGATCAGAGAACTA	1967
Db	1475	LysTyrValGlnGluLeuAlaAlaValArgAlaAspAlaGluThrArgLeuAlaGluVal	1494
QY	1968	ACGCTTGAAATCGAGAGACTGAAGAAACGGCTCCAGCAGTTGGAGGTG---GTGGAGGGG	2024
Db	1495	GlnArgGluAlaGlnSerThrAlaArgGluLeuGluValMetThrAlaLysTyrGluGly	1514
QY	2025	GACTTGATGAAGACCGAGGACGAATATGACCAGTTGGAGAGAAAGTTTCAGAACCGAGCAG	2084
Db	1515	AlaLysValLysValLeuGluGlu-----ArgGlnArgPheGlnGluGluArg	1530
QY	2085	GATAAGGCCAAACTTCCTCTCCAGCAGCTCGAGGAAATCAAACACCAAATGGCCCAAGCAC	2144
Db	1531	GlnLys-----LeuThrAlaGlnValGluGluLeuSerLysLysLeuAlaAspSer	1547
QY	2145	AAAGCCATAGAGAAAGGGAGGCCGTGAGCCAGGAAGCCGAACCTGCGACACAGGTTTCGG	2204
Db	1548	-----AspGlnAlaSerLysValGlnGln-----LysLeuLys	1559
QY	2205	CTGGAGGAGGCTAAAGTCGTGATTTACAGGCCGAGGTGCAGGCTCTCAAGGAGAAGATC	2264
Db	1560	AlaValGlnAlaGlnGlyGlyGluSerGlnGlnGluAlaGlnArgPheGlnAlaGlnLeu	1579
QY	2265	CACGAGCTGATGAACAAGGAAGACCAGCTGTCTCAGTCCCAAGTCGACTATTTCGGTCCTT	2324
Db	1580	AsnGluLeuGlnAlaGlnLeuSerGlnLysGluGlnAlaAlaGluHisTyrLysLeuGln	1599
QY	2325	CAGCAAGATTTATGGAAGAAGAAACTAAGAACAAAGATGGGAGGAGGTCTCTCAAT	2384
Db	1600	MetGluLysAlaLysThrHisTyrAspAlaLysLysGlnGlnAsnGlnGluLeuGlnGlu	1619
QY	2385	CTGACCAAGGAGCTAGAGCTTTCCAAAGCGCTACAGCCGAGCTCTCAGGCCGAGTGGGAAC	2444
Db	1620	GlnLeuArgSerLeuGluGlnLeuGlnLysGluAsnLysGluLeuArgAlaGluAla---	1638
QY	2445	GGCCGAAGATGGTGGACGTGCCTGTGGCTCCACTGGGCTGCAGACC-----GAG	2495
Db	1639	-----GluArgLeuGlyHisGluLeuGlnGlnAlaGlyLeuLysThrLysGluAlaGlu	1656
QY	2496	GCGGTGTCGGGATGCTGCGGAGGAGGAGACCCCGGCTGTGTTTCATTCCGAAATCCTTC	2555
Db	1657	GlnThrCysArgHisLeuThrAlaGln-----ValArgSerLeuGlu	1670
QY	2556	CAGGAGGAAAAATCACATCATGAGTAATCTTCGACAGGTAGGCTCTGAAGAAACCCATCGAA	2615
Db	1671	AlaGlnValAlaHisAlaAspGlnGlnLeuArgAspLeuGly---LysPheGlnValAla	1689
QY	2616	CGGTCTCGGTCTCGACAGGATATCCCCAGCAGCGAATGAGCTCACCATGAGGAAGTCT	2675
Db	1690	ThrAspAlaLeuLysSerArgGluProGlnAlaLysProGlnLeuAspLeuSerIleAsp	1709
QY	2676	TGGATTCTTGGATGAGAAAAAGAGAAAAACGGTCCITCCACTCCGAGGAGAAAGGCCCC	2735
Db	1710	SerLeuAsp---LeuSerCysGluGluGlyThrProLeuSerIleThrSerLysLeuPro	1728
QY	2736	AGGCCAAAC---CAGGTTGCAGGGCACCCCGGGGAGCTGGTC-----	2774
Db	1729	ArgThrGlnProAspGlyThrSerValProGlyGluProAlaSerProIleSerGlnArg	1748
QY	2775	CTAGACCAAAAGCAGGCCAGCCCTACACATCCGTGTGACACCAGATCATGAGAACAGC	2834
Db	1749	LeuProProLysValGluSerLeuGluSerLeuTyrPheThrProIleProAlaArgSer	1768
QY	2835	ACTGCCACCTGGAGATCACAAAGCCCAACATCTGAAGAGTTTTC-----	2879
Db	1769	GlnAlaProLeuGluSerSerLeuAspSerLeuGlyAspValPheLeuAspSerGlyArg	1788
QY	2880	-----TCTAGTACCACCGTCATTCTACCTTAGGCAACCCAGAAA	2918
Db	1789	LysThrArgSerAlaArgArgThrThrGlnIleIleAsnIleThrMetThrLysLys	1808

QY	2919	CCAAGAAATAACCATATTATCCATCACCCCAATGTCTATGTCGAAAAAGCCCAAAAGTGCAGAT	2978
Db	1809	LeuAspValGluGluProAspSerAlaAsnSerSerPheTyrSerThrArgSerAlaPro	1828
QY	2979	CCTACTCTCGGCCCCAGAACGAGCCATGTCCCTGTACGATTACTACTATTTCCAGAGAG	3038
Db	1829	AlaSerGlnAlaSerLeuArgAlaThrSer-----SerThrGlnSerLeuAlaArgLeu	1846
QY	3039	AAGAGCCCGGAAGGTGGAAGGAGCGCCTTTGCCGACAGGCCTGCA---TCCCCCATCCAA	3095
Db	1847	GlySerProAspTyrGlyAsnSerAlaLeuLeuSerLeuProGlyTyrArgPro-----	1864
QY	3096	ATCATGACGGTGTCAACATCTGCAGCTCCCACACTGAAATCGCTGTCTCTCTGAATCTCAG	3155
Db	1865	-----ThrThrArgSerSerAlaArgArgSerGlnAlaGlyVal-----SerSer	1879
QY	3156	GAAGTCGCTATGGGAAGGACTATCTCTAAAAGTCAACCCCGGAAAAACAACACTGTTCCAGCC	3215
Db	1880	GlyAlaProProGlyArgAsnSerPheTyrMet---GlyThrCysGlnAspGluProGlu	1898
QY	3216	CCCGTCGGGAAGTACAACTCCAATGCTAAATATATCATCACCCGGAAGACAATAAATT---	3272
Db	1899	GlnLeuAspAspTIpAsnArgIleAlaGluLeu-----GlnGlnArgAsnArgValCys	1916
QY	3273	-----CACATTCACTGGGTTCTCAGTTTAAAGCGATCTCTGGCCTGCCGTGAAGGC	3326
Db	1917	ProProHisLeuLysThrCysTyrProLeuGluSerArgProSerLeuSerLeuGlyThr	1936
QY	3327	GTGAGC-----CCAGTTATCACCGTCCGG-----	3350
Db	1937	IleThrAspGluGluMetLysThrGlyAspProGlnGluThrLeuArgArgAlaSerMet	1956
QY	3351	--CCTGTCAACGTGACAGCGGAGAAGGAGGTTTCTACAGGCACAGTCCTTCGCTCTCCC	3407
Db	1957	GlnProIleGlnIle---AlaGlu-----GlyThrGlyIleThrThrArgGlnGln	1972
QY	3408	AGGAACCACTCTCT------TCAAGACCCCGTGTAGCAAAAGTG	3446
Db	1973	ArgLysArgValSerLeuGluProHisGlnGlyProGlyThrProGluSerLysLysAla	1992
QY	3447	ACCAGCACTATA-----ACTATAACCCCG-----GTC	3473
Db	1993	ThrSerCysPheProArgProMetThrProArgAspArgHisGluGlyArgLysGlnSer	2012
QY	3474	ACAACGTATCCACACAGAGGAACCCAATCAGTGTACAGGACAAGATGGGTATCTCAGCGG	3533
Db	2013	ThrThrGluAlaGlnLysLysAlaAlaProAlaSerThrLysGlnAlaAspArgArgGln	2032
QY	3534	CCT-----ACCCCCACCCGCATTCTATGTCAAAAGGTATGAAA	3572
Db	2033	SerMetAlaPheSerIleLeuAsnThrProLysLysLeuGlyAsnSerLeu---LeuArg	2051
QY	3573	GCTGGAAGCCAGTAGTGGCAGCCTCAGGAGCAGGAAATCTGACCAAAATTCAGCCTCGA	3632
Db	2052	ArgGlyAlaSerLysLysAlaLeuSerLysAlaSerProAsnThrArgSerGlyThrArg	2071
QY	3633	GCTGAGACTCAGTCTATGAAAAATAGAGCTGAAGAAATCTGCAGCCAGCAGCACTGCCTCT	3692
Db	2072	-----ArgSerProArgIleAlaThrThrAlaSerAlaAlaThrAlaAlaAla	2088
QY	3693	CTTGA 3698	
Db	2089	IleGly 2090	

Search completed: September 7, 2004, 14:12:00
Job time : 353.5 secs